GAKUTO International Series
Math. Sci. Appl., Vol. 31 (2009)
Multiple scales problems in Biomathematics, Mechanics, Physics and Numerics, pp. 133–181

GAKKOTOSHO
TOKYO JAPAN

The Finite Element Heterogeneous Multiscale Method: a computational strategy for multiscale PDEs

A. Abdulle

Abstract: Heterogeneous multiscale methods (HMM) have been introduced by E and Engquist [*Commun. Math. Sci. 1 (2003), pp. 87-132*] as a general methodology for the numerical computation of problems with multiple scales. In this paper we discuss finite element methods based on the HMM for multiscale partial differential equations (PDEs). We give numerous examples of such multiscale problems, including elliptic, parabolic and advection diffusion problems and discuss several applications in areas such as porous media flow, biology and material sciences. A detailed analysis of the methods as well as recent developments are discussed.

1. Introduction

Multiscale or multi-physics modeling play a major role in many important problems arising in the geosciences, atmospheric sciences, physical sciences, chemistry, biology or medicine. Without attempting to be exhaustive we mention the following: the study of groundwater pollution through infiltration of a fluid in a porous medium, the understanding of the effect of subgrid processes such as convection and cloud formation in climate modeling, the effective properties of composite materials increasingly used in engineering, the simulation of chemical processes mixing particles whose concentrations differ from several orders of magnitude, the mechanical properties of heterogeneous tissues as bones, important to understand mechanisms which lead to crack, failure or diseases.

Increasing capabilities in experimental sciences and new questions on the interactions of the fundamental building blocks of nature have raised major theoretical and computational problems and urged for the development of multiscale mathematics. The past few years have

seen increasingly growing research activities aiming at developing novel multiscale computational methods. While traditional approaches were based on sequential strategies with empirical macroscopic models derived with parameters computed beforehand from microscopic models, new methods based on simultaneous coupling or "on-the-fly computations", extracting coarse dynamics from multiscale systems have emerged [20],[32],[34],[39].

In this paper we focus on multiscale problems modeled as partial differential equations (PDEs) belonging to the class of so-called homogenization problems. Analytic treatments of such problems have been studied for many years [22],[26],[45] and are still an active field of research. Homogenization theory is concerned with the macroscopic description of a microscopically heterogeneous system. The impacts of the small scales of such systems at a macroscale are usually non-trivial and finding the "right" averaging process is at the heart of homogenization theory. The advantages of considering a homogeneous system by averaging out the fine scales are twofold: first, it simplifies the understanding of the macro dynamics of the considered problem; second, it reduces considerably the cost of numerical simulations. In many cases a computational approach of a full system with complex scale interactions is out of reach, even with nowadays powerful supercomputers. These computational and modeling issues have triggered the recent development of numerical methodology for multiscale (homogenization) problems.[1] Among them, the so-called heterogeneous multiscale method (HMM) has proved to be an efficient tool to assemble information from microscale problems in order to perform macroscale simulations. These methods introduced by E and Engquist have already been used successfully in several applications [34] and are still under active developments. In this paper we discuss the modeling and analysis of finite element methods (FEMs) for multiscale problems constructed in the framework of the HMM for multiscale PDEs. The HMM strategy, as we will see, offers many advantages:

- it works for different type of problems and operators,
- it allows for algorithms which are not restricted to specific assumptions on the small scales,
- it allows flexibility in the type of discretization,
- it offers a good framework for analysis and implementation.

Many other approaches, often (but not exclusively) tailored to elliptic problems have been developed. It is not our intention to review them and we just mention here a few references. Numerical computations for homogenization problems was pioneered by Babuška [19] for elliptic problems and Engquist [30] for dynamic problems. Dorobantu, Engquist and Runborg [29],[31] proposed a method based on multi-resolution analysis, Neuss, Jäger and Wittum [58] proposed a method based based on multigrid with homogenization used in the coarsening process, Hou and co-workers proposed the multiscale finite element method (MSFEM) based on modified basis functions obtained from the fine scale equations [35],[42], Babuška, Matache and Schwab developed the two-scale FEM [52],[53], Viet Ha Hoang and Schwab proposed the high dimensional FEM [44]. For a description of the pros and cons

[1] We note that in the structural mechanics or engineering communities, one often use the terminology of "representative volume element" (RVE) for such averaging processes, while in the porous media or physics communities one often refers to "upscaling".

of these techniques and a comparison with HMM, we refer to [34] and [54]. We also notice that there is a huge literature concerned with micro-macro methods based on representative volume elements (RVEs) in the structural mechanics and engineering communities. The methods have been proposed for various type of problems, however often without convergence analysis. We mention Terada, Kikuchi and co-workers [63], Kouznetsova, Baaijens and co-workers [48] and Miehe and co-workers [51].

In this paper we discuss recent developments of the HMM for the modeling analysis and computation of multiscale PDEs. The discussion is organized as follows. We start in Section 2 by presenting several examples of multiscale problems and their simulation with the HMM. In Section 3 we discuss in detail the modeling and the analysis of the finite element heterogeneous multiscale method (FE-HMM). In Section 4, we present some recent developments of the FE-HMM for PDEs, as hybrid methods coupling spectral or discontinuous Galerkin methods with FEM. Finally, we conclude with some remarks about some issues and new directions of research to enhance the computational capabilities of the FE-HMM.

Notations. In what follows, $C > 0$ denotes a generic constant, independent of ε, whose value can change at any occurrence but depends only on the quantities which are indicated explicitely. For $r = (r_1, \ldots, r_d) \in \mathbb{N}^d$, we denote $|r| = r_1 + \ldots + r_d$, $D^r = \partial_1^{r_1} \ldots \partial_d^{r_d}$. We will consider the usual Sobolev space $H^1(\Omega) = \{u \in L^2(\Omega); D^r u \in L^2(\Omega),\ |r| \leq 1\}$, with norm $\|u\|_{H^1(\Omega)} = (\sum_{|r|\leq 1} \|D^r u\|^2_{L^2(\Omega)})^{1/2}$. We will also consider $H^1_0(\Omega)$ the closure of $C^\infty_0(\Omega)$ for the $\|\cdot\|_{H^1(\Omega)}$ norm and the spaces $W^{l,\infty}(\Omega) = \{u \in L^\infty(\Omega); D^r u \in L^\infty(\Omega),\ |r| \leq l\}$. Finally for the unit cube $Y = (0,1)^d$, we will consider $W^1_{per}(Y) = \{v \in H^1_{per}(Y);\ \int_Y v dx = 0\}$, where $H^1_{per}(Y)$ is defined as the closure of $\mathcal{C}^\infty_{per}(Y)$ (the subset of $\mathcal{C}^\infty(\mathbb{R}^d)$ of periodic functions in Y). Finally, we will use the notation $|\cdot|$ for the standard Euclidean norm in $\mathbb{R}^d$.

2. Computational strategy and examples

In this section we present examples of multiscale problems arising in various applications. Although the problems originate from very different fields of research, there is, as we will see, a common strategy to model and discretize them. We first discuss briefly the type of multiscale problems we will consider, their analytic treatments and the methodology of the HMM.

2.1 Computational strategy

We are interested in solving PDEs in a computational domain Ω with coefficients originating from some fine scale structure. We write such problems as

$$\mathcal{L}^\varepsilon(u^\varepsilon) = f^\varepsilon,$$

where $\mathcal{L}^\varepsilon$ denotes some differential operator, u^ε some quantity of interest and f^ε some data of the problem. Here and in what follows, ε represents one or several microscopic scales of the problem (that we assume to be well separated) and we assume that a macroscopic description (at least in part of the computational domain) exists. We are interested in situation where the solution u^ε is required (at first approximation) only in some averaged sense. A typical example arises with composite materials when two or more materials are finely mixed together. At a micro scale, we have small heterogeneities (that we suppose distributed with some self-similarity) and the thermal conductivity of the body oscillates between the

thermal conductivities of its constituents. The small variations in the thermal properties are usually not the primal interest, but one would rather like to know the "effective" property of the composite, i.e., when observed at a larger scale at which it looks "homogeneous". The question is thus to understand the macroscopic dynamics of systems governed by microscopic heterogeneities.

Macroscopic dynamics. The class of problems we have in mind can have many scales, but a crucial assumption is that of scale separation. This is realistic for many applications although sometimes only in some region of the computational domain and/or for some period of time (see Section 5 for discussions on this issue). The assumption of scale separation allows to use mathematical tools such as averaging methods or homogenization/perturbation theory, describing the effective dynamics of the aforementioned multiscale problems. In a PDE context, we will focus on homogenization theory which describes the macro dynamics of systems governed by microscopic scales. Let us first reformulate this problem in the following way: can we replace a given heterogeneous medium by a homogeneous medium with similar large scale properties ? At the mathematical level, homogenization is concerned with finding a limit solution u^0 for u^ε when $\varepsilon \to 0$ and an equation for it

$$\mathcal{L}^0(u^0) = f^0.$$

The solution u^0 obtained as a limit (to be made precise later) of u^ε does no longer depend on the small scale ε and is called the homogenized solution. Back to our physical problem, it can for example describe the temperature distribution in an ideal composite material, in which the phases are perfectly mixed and that is thus homogeneous.

Numerical issues. The problems arising in computational approaches of such problems are the following. On one hand, the computational cost associated with the discretization of $\mathcal{L}^\varepsilon(u^\varepsilon) = f^\varepsilon$ is usually very high, since with any standard method one needs to resolve the small scale of length ε of the problem and implement the method with a meshsize $h < \varepsilon$. If ε is small, this approach is often not feasible. On the other hand, the equations for the homogenized problem are usually not available in closed form.

HMM methodology. The methodology of the HMM can be summarized as follows.

- Step 1 (modeling): define a macroscopic discretization with *macroscopic* input data recovered by averaging on the fly microscopic simulations obtained from the available *microscopic* problem.
- Step 2 (computation): extract a macroscopic solution based on the macro to micro modeling.
- Step 3 (post-processing and adaptivity): recover the fine scale information and/or refine the macroscopic discretization where needed.

The two first steps are based on a scale separation assumption, averaging theorems (as homogenization) and the physics at the macroscopic level for the appropriate modeling (for example conservation laws). The third step is a post-processing process or an adaptive procedure based on the specific problem (scale separation may only be valid in part of the computational domain or for a given time of a dynamical process).

2.2 Examples

We describe here several examples from various applications, discuss their multiscale modeling and explain how a common computational strategy can be developed for their numerical solutions.

2.2.1 Macromolecules transport in microfabricated sieve. The problem of separating large biomolecules such as DNA is fundamental for biological research and biomedical applications. The main technique currently used is the separation through gel electrophoresis. The process works as follows: DNA fragments (whose phospate backbone are negatively charged) are placed into a device filled up with a porous gel (usually agarose) and an electric field forces the fragments to migrate through the gel (see Figure 1). The macromolecules are

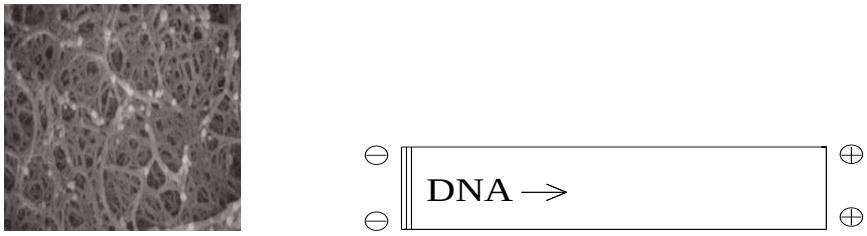


Figure 1: Snapshot of Agarose gel at μm scale (left picture), separation device filled with the gel (right picture).

then "sieved" and separated in a size-dependent manner thanks to the porose structure of the gel. Although the technique of choice for separation of macromolecules, gel electrophoresis has several drawbacks as the duration of the process (up to several hours) and the cost (each new fragment needs a new gel matrix). There has thus been an increasing interest over the last few years to find alternative ways to achieve separation.

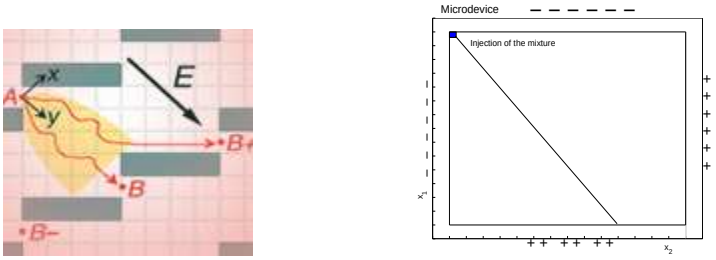


Figure 2: Asymmetric obstacle arrays of μm size (left picture), sorting device with asymmetric obstacles (right picture), left picture from [28, p.155]. The injection area is at the upper left corner, the collection of the fragments at the lower right corner of the device (left picture). According to [28], the trajectory of the smaller molecules should deviate more from the diagonal direction than the trajectory of the larger ones.

Inspired by "lab-on-chip" ideas, Duke and Austin [28] and Ertas [36] proposed to replace the porous gel by a solid microstructure composed of asymmetric obstacles of micrometer length, and let the DNA fragments migrate in such a device due to the forces of an electrical field (see Figure 2). It was believed that diffusive motion combined with the special asymmetry would deflect smaller molecules with higher diffusivity from the mean direction.

Thus, a "size-based" separation of macromolecules could be obtained (see Figure 2). However, the experimental setup in [41] showed that particles much smaller than the barrier gap of the obstacles are poorly separated. In [3] a closed theory (based on homogenization theory) was given to explain the experimental findings. Furthermore, in [3] and [4] numerical multiscale techniques were developed to simulate such an experimental setup. The ability to test numerically such prototype or other "lab-on-chip" devices before construction is of high interest. The challenge in the numerical simulation of transport processes in such devices usually resides in the multiscale nature of its components (spatial heterogeneities from nm to cm) as well as in the various time scales of the dynamics.

Modeling. According to Ohm's law the flux of electrically charged particles J_E is given by $J_E = c\mu E$ where c denotes the particle concentration and μ the mobility. The mobility μ is related to the electrical conductivity a by $\mu = \rho a$, where ρ is the charge density of the electrical array which we assume to be constant for simplicity and set to one. In contrast to standard electrophoresis where separation is achieved due to particle size dependent mobility, here the mobility is assumed to depend only on the geometry of the microarray. Defining a velocity field $v = \mu E = aE = a\nabla u$, where u is the electrical potential, we obtain

$$\nabla \cdot (a\nabla u) = 0, \tag{1}$$

with Dirichlet and Neumann boundary conditions (see Figure 2 right picture, the Neumann boundary conditions are at the (insulated) corner and the Dirichlet boundary conditions at the remaining parts of the boundary). To obtain the total particle flux, a diffusive flux is added

$$J = c\mu E - D\nabla c, \tag{2}$$

so that the mass conservation law for the particle concentration reads

$$\frac{\partial c}{\partial t} + \nabla \cdot J = 0. \tag{3}$$

Two typical length scales are present in the above problem: a microscopic length scale l (of size μm) representing the size of the obstacles, and a macroscopic length scale L (of size cm) at which the transport behavior is observed. The asymmetric obstacles induce a typical microscopic self-similarity (see shaded area in Figure 3) and we set ε to be its length scale. This parameter is obviously proportional to l/L. Thus, the conductivity tensor in (1) will depends on ε and we denote it by $a^\varepsilon(x)$. As a consequence, the potential u, solution of (1) as well as the velocity field and the concentration in (3) will depend on ε. By rescaling the equation (1) and (3) according to the micro and macro length scales we obtain the following system of multiscale equations

$$\begin{aligned} \nabla \cdot (a^\varepsilon \nabla u^\varepsilon) &= 0, & (4)\\ \frac{\partial c^\varepsilon}{\partial t} + v^\varepsilon \nabla c^\varepsilon &= \nabla \cdot D\nabla c^\varepsilon, & (5) \end{aligned}$$

with suitable initial and boundary conditions. As mentioned earlier in this section, the question is now to understand the macroscopic dynamics (see Figure 2 right picture) from the system (4-5) governed by microscopic heterogeneities (see Figure 2 left picture).

Analytically, homogenization theory is the right tool to derive (non explicit) macroscopic dynamics. The formal derivation is obtained by a multiple scale expansion of the concentration

$$c^\varepsilon(t,x) = c_0(t,x) + \varepsilon c_1(t,x,x/\varepsilon) + \varepsilon^2 c_2(t,x,x/\varepsilon) + \dots,$$

where the first term c_0 will be identified with the homogenized solution. Here x is the slow scale and x/ε the fast (oscillating) scale. Likewise, the (divergence-free) velocity field v^ε is splitted into a large scale component v_0 and a fluctuating (zero-mean) component $\tilde{v}$ as $v^\varepsilon(x) = \bar{v}(x) + \tilde{v}(x, x/\varepsilon)$. Inserting the asymptotic expansion for c^ε and the splited velocity field in (5) and identifying the power of ε, we obtain a cascade of equations from which we can deduce the homogenized equation

$$\frac{\partial c_0}{\partial t} = \bar{v}\nabla c_0 + \nabla \cdot D_0 \nabla c_0, \tag{6}$$

where D_0 is an enhanced effective diffusion tensor [3],[49],[60]. The mean velocity field $\bar{v}$ can then be approximated by $\bar{v} = -a^0(x)\nabla u^0 + \mathcal{O}(\varepsilon)$, where u^0 is the solution of a homogenized elliptic problem

$$-\nabla \cdot \left(a^0 \nabla u^0\right) \;=\; 0, \tag{7}$$

where $a^0(x)$ is the so-called homogenized conductivity tensor [3],[7]. Notice that the tensor $a^0(x)$ is usually not available in explicit form and its computation relies on the solution of elliptic problems, the so-called cell problems (see Section 3). Such equations have in theory to be solved for each point x of the domain which is of course impossible in practice and one has to localize its computation.

Thanks to the homogenization process, the heterogeneous fine scale model (5) is transformed into a homogeneous large scale model (6) which describes the macroscopic behavior of the transport of the particles. We see in equation (6) that particles with different molecular weights (or diffusion constants) will move with the same direction given by the effective drift. Thus, for particle transport in a heterogeneous divergence-free flow field, no diffusion dependent deflection of particles from the mean flow direction take place and no particle separation can occur. This explains the experimental finding presented in [41]. Let us remark that such effects (e.g. trapping phenomena) can be obtained with non-divergence free flow fields. This has been recently studied in [16].

Numerical experiments. Even though very useful to understand the macroscopic dynamics, analytical techniques such as homogenization are not explicit enough to allow for practical computations of transport phenomena in heterogenous media and the use of numerical methods is required. However, the applicability of standard numerical techniques is not obvious as the discretization of equations (4-5) leads to a problem of enormous dimension if the size of the obstacle are much smaller than the size of the device. Indeed for such techniques, the meshsize h used in the calculations must usually be smaller than the microscopic structure, i.e. $h < \varepsilon$. As explained in Section 3, the finite element heterogeneous multiscale method (FE-HMM) allows to compute approximations $u^H, v^H = a^0(x)\nabla u^H$ of the homogenized potential and velocity fields $u^0, \bar{v}$, respectively, where the superscript H refers to a typical meshsize used in numerical computations (here $H >> \varepsilon$ is allowed). The dynamics for the fine scale concentration c^ε is sometimes also of interest. Through a post-processing process, an approximation of the fine scale velocity field v^ε can be obtained with

the FE-HMM, allowing for a computation of c^ε. Finally the numerical value of the tensor $a^0(x)$, the homogenized conductivity, can be obtained with the in the FE-HMM if required, as a by-product, during the computation of u^H.

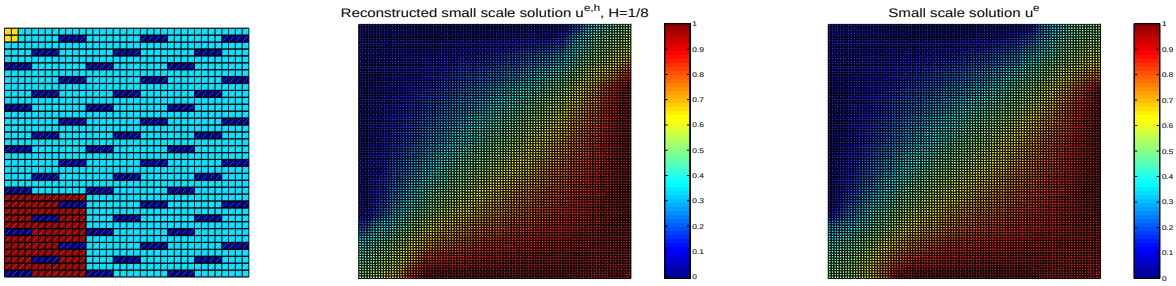


Figure 3: Snapshot of a device with μm obstacles, with in shaded area a typical self-similarity of the asymmetric structure (left picture), solution of the multiscale problem for the electrical potential (4) (middle and right picture). Middle picture is the FE-HMM solution with about 10^3 DOF, right picture is a reference solution computed on a fine grid using about 10^6 DOF.

To solve the transport problem, we have first to handle the elliptic problem (4) to obtain a mean or fine-scale velocity field. In Figure 3 (middle and right pictures), we compare a reference (resolved) solution for the fine scale potential u^ε and its numerical approximation $u^{\varepsilon,h}$ obtained by the FE-HMM through a post-processing process. The domain is scaled to be the unit square and we apply Neumann boundary conditions at the corner (insulated region) and Dirichlet boundary conditions on the sides of the domain (applied electric potential) (see Figure 4, left picture). The value of ε is chosen to be $\sim 10^{-3}$. The middle picture in Figure 3 is obtained with the FE-HMM strategy on a coarse meshsize $H = 1/8$ (with post-processing). The computation with the FE-HMM involves about 10^3 degrees of freedom (DOF). As explained in Section 3, the DOF do not depend on the size of the small scale for a problem with self-similarity and scale separation. It does only depend on the macro mesh and the number of sampling points of the microstructure. In the right picture of Figure 3 we sketched a reference solution for the problem (4), involving about 10^6 degrees of freedom. We see that we have a good agreement between both solutions. Let us emphasize that for the reference solution, the complexity depends on ε. With $\sim 10^{-5}$ we face a problem of about 10^{10} DOF for the fine scale solution, impossible to solve, while the FE-HMM strategy will still only need about 10^3 DOF (for the same quality of approximation as with $\varepsilon \sim 10^{-3}$).

Once we have a numerical approximation of the fine scale velocity field, we can solve the transport problem. For that, we use the method of lines (discretization of the spatial variables only) to obtain a system of ordinary differential equations (ODEs) which has to be solved by an appropriate ODE solver. Since a fine mesh is needed (if we want to compute the fine scale transport problem), it will lead to a problem of large dimension. Furthermore, the problem is stiff, which means that many time scales are involved in the dynamics. This makes standard explicit methods (as the Euler method) inefficient since the time step is constrained by the fastest time scale involved in the problem. The usual wisdom in such situations is to use an implicit solver. But the drawback with such an approach is the requirement to solve nonlinear systems (which can be large as in the present problem) at each time step. Here, we opt for a good compromise: the ROCK method. This method belongs to the so-called class

of Chebyshev methods and exploit stabilization techniques obtained through Chebyshev-like stability polynomials to allow for much larger time steps in stiff computations [1]. At the same time, the ROCK methods are explicit and thus as simple to use as the Euler explicit method.

We compare in Figure 4 the evolution of the particles advected by a reference velocity field obtained via scale resolution (right picture) and a reconstructed velocity field (middle picture) obtained with the FE-HMM (with post-processing). The spatial discretization for the transport problem is the same in both experiments. We perform the time integration for $t \in [0, 1.2]$ and record the solution at discrete time $t = 0, 0.3, 0.6, 0.9, 1.2$, to compare the evolution of the two transport problems. We see in Figure 4 that the dynamics with the reconstructed velocity field (obtained with the FE-HMM) agrees very well with the dynamics depending on the fine scale velocity field, illustrating the efficiency of the proposed numerical method. Details and additional numerical experiments for such problems can be found in [3],[7],[4].

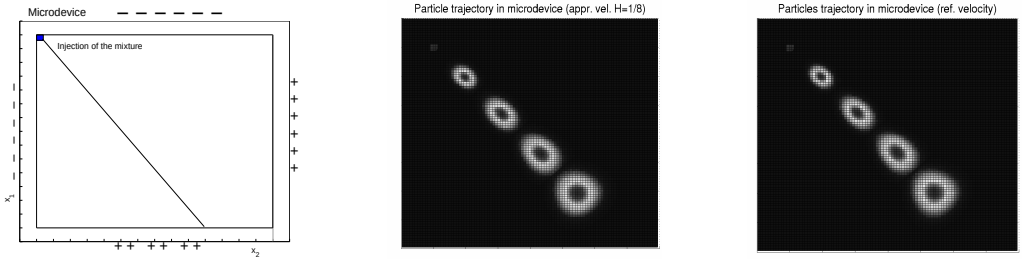


Figure 4: Computational domain (left picture). The Neumann boundary conditions are on the four corners and the Dirichlet boundary conditions everywhere else. Middle and right pictures: simulation of the particles trajectory starting from the upper left corner. Each cloud of particles represent its location at a given time. The velocity field is obtained from the FE-HMM (middle picture) while a reference fine scale velocity field is used in the right picture.

2.2.2 Water infiltration in porous medium. A basic problem in hydrology and soil physics is that of absorption of water in a porous medium. Understanding this process is important for water resource management and the understanding of environmental problems caused for example by underground pollution. A widely used model to describe flow of water in unsaturated porous media has been proposed by Richards in 1931.

Modeling. To describe Richards' model, we start with a mass balance equation

$$\frac{\partial \Theta}{\partial t} + \nabla \cdot q = 0, \tag{8}$$

where Θ is the water content and q the water flux and where we neglected source and sink terms for simplicity. For saturated medium, a well-known relation between the water flux q and the fluid pressure u is given by the Darcy law $q = -a\nabla u$, where a is the conductivity tensor. For unsaturated media, the conductivity will depend on the water content and the above relation reads $q = -a(\Theta)\nabla u$. In view of (8) we obtain the Richards equation

$$\frac{\partial \Theta(u^\varepsilon)}{\partial t} - \nabla \cdot \Big(a\big(\Theta(u^\varepsilon)\big)[\nabla u^\varepsilon + z)] \Big) = 0, \tag{9}$$

where $z = -\rho g$ represents an additional term due to the influence of gravity and where ρ is the water density and g the gravitational acceleration. The problem (9) becomes a multiscale problem if one takes into account the dependence of the infiltration process upon the heterogeneity of the medium. Indeed, the conductivity a can vary locally at a much smaller scale (pore scale) denoted here by ε, than the scale of observation. To emphasize on these multiscale effects, we add a superscript ε to the conductivity and the pressure in the problem (9). We notice that (9) is a nonlinear equation which can furthermore degenerate, from parabolic to elliptic, when the medium becomes saturated [62].

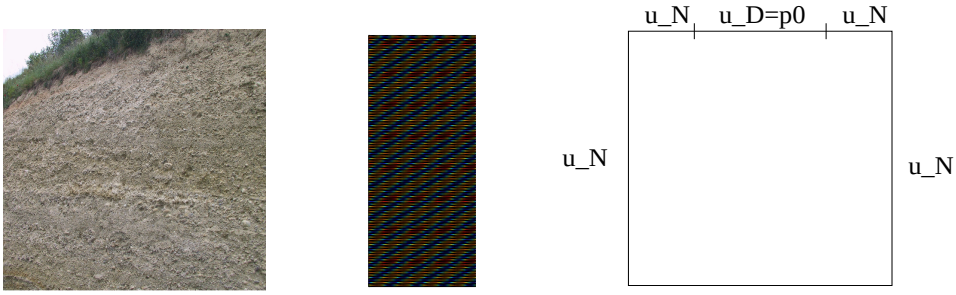


Figure 5: Real heterogeneous soil (left picture), snapshot of a model for layered medium (middle picture), computational domain with Neumann u_N and Dirichlet u_D boundary conditions (right picture).

Homogenization is again the right analytical tool to derive a macroscopic dynamics for the equation (9) and to address such questions as the existence of limit solutions $u^\varepsilon \to u^0$, $\Theta(u^\varepsilon) \to \Theta(u^0)$ and the existence of a limit equation of the type (9) for these quantities. Homogenization of nonlinear equations similar to (9) has been studied recently and we refer to [50],[43],[13] and the references therein.

Numerical experiments. Numerous methods have been proposed for the numerical solution of the Richards equation. Without attempting to be exhaustive, we mention [61] and [62] and the references therein. The methods in [61],[62] are fine scale approaches and aim at solving the original fine scale equation (9). The issue of degeneracy of Richard equation is also addressed in these papers. Much less work has been done within a multiscale approach in which we usually want to recover the effective behaviour of the system without solving all its fine scale details. We mention [35] where a numerical strategy based on the so-called multiscale finite element method (MsFEM) has been proposed and [27] where the nonlinear constitutive relation are upscaled before solving the nonlinear problem. In [13] we proposed a numerical method based on HMM, where coarse graining and macro discretization are performed simultaneously, allowing a substantial saving in terms of computational cost compared to the full fine scale solution of the original equation. The time integration uses a linearization process first described in [62].

We describe briefly a numerical simulation for an infiltration problem. The numerical method used here was first proposed in [13] and is inspired from the FE-HMM. Notice that the nonlinearity and the time dependence have to be properly addressed and we refer to [13] for details. To solve problem (9), constitutive relations for $\Theta(u^\varepsilon)$ and $a(\Theta(u^\varepsilon))$ are needed. Among many models, empirical formulations of these constitutive relations are given by

the Haverkamp, the van Genuchten and the exponential models (see [37] and the references therein). Here, we use an exponential constitutive relation given by $\Theta(u) = \Theta_s e^{\beta u}$, where $\Theta_s = 1$ is the saturated water content (we choose $\beta = 0.1$ in the simulations below). The fluctuations are modeled through the conductivity as $a(\Theta(u^\varepsilon)) = k_\varepsilon e^{\alpha_\varepsilon u}$, with

$$k_\varepsilon(x) = \frac{c}{2 + 1.8\sin(2\pi(2x_2 - x_1)/\varepsilon)}, \tag{10}$$

$$\alpha_\varepsilon(x) = 10k_s^\varepsilon(x), \tag{11}$$

where $\varepsilon = 1/16$, the constant $c = 1/114.7$ is chosen such that $\bar{a} = \int_Y a(y) = 0.01$ and $\bar{k} = \int_Y k(y) = 0.1$ (these parameters are borrowed from [27] and [35] and allow to compare numerical results). The layered permeability field $k_\varepsilon(x)$ is depicted in Figure 5 (middle picture). The computational domain is shown in Figure 5 (right picture), where u_N, u_D refer

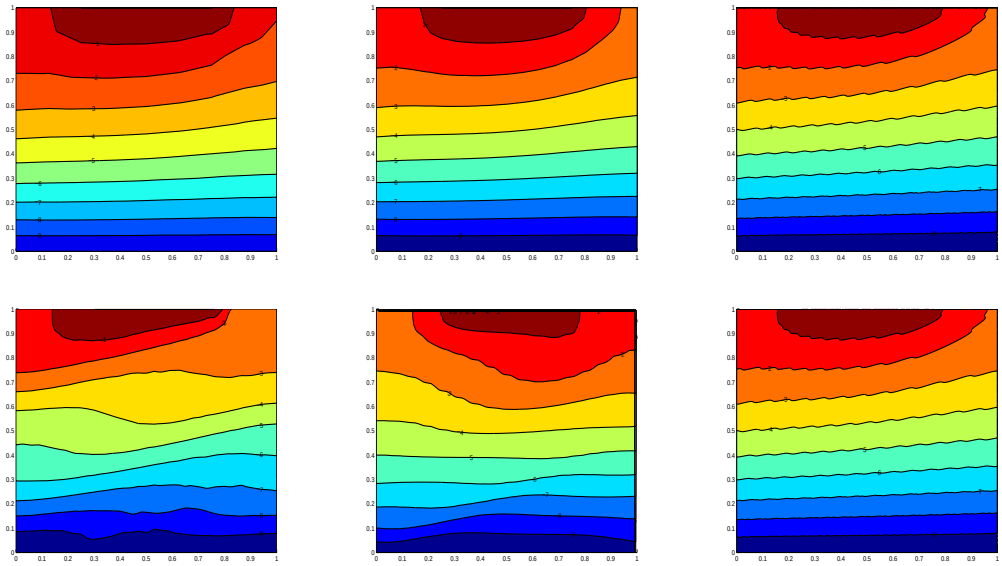


Figure 6: First row of pictures: level curves for the pressure u for problem (9) on coarse meshes 8×8 and 16×16 (first and second pictures) with the FE-HMM like method; the last picture is a reference solution. Second row of pictures: computations on the same coarse grid with a standard method; the last picture is again the reference solution.

to Neumann and Dirichlet boundary conditions, respectively. A constant initial condition $u^\varepsilon(x, 0) = u_0$ is given and we solve the problem (9) over the time interval $[0, 10]$ with a time step $\Delta t = 0.5$. For the FE-HMM inspired method, we choose successively two coarse meshes of size 8×8 and 16×16 (the computational domain is scaled to $(0, 1)^2$). We monitor at $T = 10$ the level curve of the pressure and compare it for each coarse mesh to a reference solution obtained by solving the original equation on a fine grid, resolving the heterogeneities. We can see in Figure 6 (first row of pictures) that the results of the multiscale method are in good qualitative agreement with the reference solution. Since $u^\varepsilon \to u^0$ strongly in the L^2 norm, the comparison with the fine scale solution gives insight into the behavior of the proposed numerical method. As explained in the introduction and discussed in Section 3, a fine scale numerical solution can be obtained with our multiscale method from the known coarse solution by a post-processing process. Finally, in Figure 6 (second row of pictures) we

give the results of a standard solver on the same coarse meshes as before (8×8 and 16×16) to illustrate that such methods are not able to capture the right infiltration process if the fine scales are not properly resolved.

2.2.3 Heat dissipation in composite materials. Composite materials, i.e. engineered materials, made from two or more material constituent have a long history. Early examples, as composite materials made of straw and mud in the form of bricks for building construction, go back to the antiquity. The ability of composite materials to have significantly different properties than its constituents makes them very attractive for optimizing material performance in a variety of applications. We mention the use of composite materials in medicine (new biomaterials for implants) and space science (carbon composite material for spacecrafts), to stress only on two very different areas of applications. Here we present yet another application related to the use of such materials in microelectronics. The use of new composite materials offers novel possibilities for chip design to cope with the development of increasingly smaller electronic components. A central issue in microprocessors developments is the ability to control the cooling process. Composite materials are routinely used in chip design as for example for the leadframe supporting the die (small block of semiconducting material on which a given functional circuit is fabricated) and for the heatsink used for the cooling process (see the references in [12]). Recently, promising attempts to use carbon nanotubes in the cooling process have also been reported [47]. The ability to test numerically

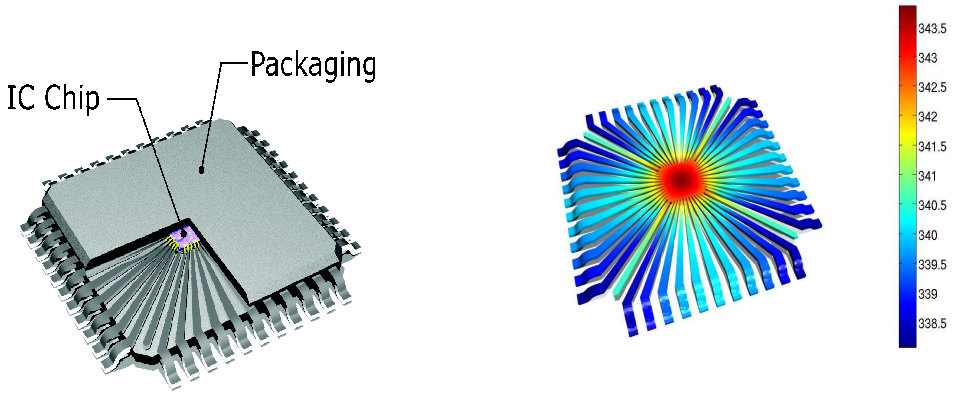


Figure 7: Leadframe and package of an integrated circuit (IC) chip (left picture), temperature distribution in the leadframe (FE-HMM solution of problem (12) with $\varepsilon = 10^{-5}$)

the properties of novel composite materials used within a microprocessor is an invaluable help for the design of new microchips and can avoid to construct costly prototypes at an early stage of development. The huge scale gap between the heterogeneities of the composite materials (from nm to μm) compared to the size of the chip (cm) makes simulation very challenging. In the following example we simulate the heat distribution in a leadframe due to the activity of the die. We only consider the metal wireframe and the IC chip, ignoring the plastic or ceramic package leading to a body as shown in Figure 7. Simulations including the packages as well as other experiments including heat distribution in heatsinks can be found in [12].

Modeling. The equations for the heat distribution in the IC chip based on the Fourier's

law of cooling are given by

$$\begin{aligned} -\nabla \cdot (a\nabla u^\varepsilon) = &\quad f, &&\text{in } \Omega, \\ n \cdot (a\nabla u^\varepsilon) = &\quad g_N &&\text{on } \partial\Omega_N, \\ n \cdot (a\nabla u^\varepsilon) + c_R u = &\quad \alpha(T - u^\varepsilon) &&\text{on } \partial\Omega_R, \end{aligned} \tag{12}$$

where $\Omega \in \mathbb{R}^3$ is the whole domain, $\partial\Omega_N$ is the surface of the chip and $\partial\Omega_R$ is the surface of the wires. The heat source, originating from the activity of the die, is modeled by a Neumann boundary condition on $\partial\Omega_N$ (heat flux spreading through the leadframe) while the boundary condition on $\partial\Omega_R$ represent the heat exchange with the environment. A heatsink is usually build on the top of the chip on larger processor but we do not consider this situation here. We emphasize that simulation for such a device (leadframe, package and heatsink) can be done with the same multiscale method as used here [12]. Since the leadframes are usually made out of composite materials (as for example copper based alloys), the conductivity tensor in (12) will depend on the microstructure of the material and we emphasize as usual this dependency on a small scale by the parameter ε, a typical size of the self similarity of the heterogeneities.

Numerical experiments. We compute a numerical simulation of the problem (12) with the FE-HMM. The multiscale tensor is chosen as

$$a^\varepsilon\left(x, \frac{x}{\varepsilon}\right) = 125 \cdot \operatorname{diag}\left(\cos\left(2\pi\frac{x_1}{\varepsilon}\right), \cos\left(2\pi\frac{x_2}{\varepsilon}\right), \cos\left(2\pi\frac{x_3}{\varepsilon}\right)\right) + \left(125 \cdot e^{50\left(x_1^2+x_2^2\right)}\right) \cdot I_3, \tag{13}$$

modeling material properties with self-similarity (here periodicity) and non-periodic slow variation from the center of the leadframe to the periphery (see Figure 8).[2] With microstruc-

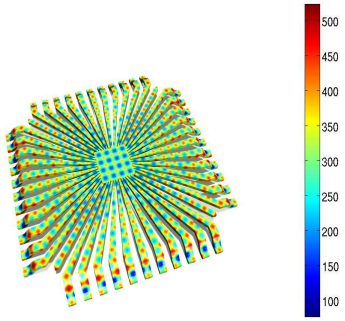


Figure 8: Magnitude for one component of the multiscale tensor ($\varepsilon = 10^{-3}$).

tures of size $\varepsilon \sim 10^{-5}$ (this arises for example with copper based alloys) a full discretization of the leadframe with a standard FEM and a mesh $h < \varepsilon$ would lead to a problem with more than 10^{10} DOF (using for example about 10 points per oscillation length), which is very hard to solve routinely. Observe that since the heterogeneous tensor is not uniformly periodic, a

[2]This model represents of course a fictitious material. Nevertheless, many composite materials exhibit such self-similarities and non-local effects. Thus, together with its nontrivial geometry, the considered model is an interesting benchmark problem to test our numerical method.

sequential strategy consisting in pre-computing the homogenized tensor by standard homogenization techniques (which must be done throughout the whole 3D domain) and solving an approximate problem of the effective heat distribution with a standard FEM, poses serious problem in terms of implementation and error control. For example, the precision and the location at which the effective tensor are precomputed will have a non-neglectible impact on the macroscopic heat distribution. The FE-HMM is capable of handling this problem with a complexity independent of the small-microstructure (assuming a self-similarity as for the present calculation). In Figure 7 (right picture) we present a simulation with the FE-HMM of problem (12) with $\varepsilon = 10^{-5}$. The simulation is done on a coarse mesh consisting of 54,000 tetrahedra with 17,000 grid points. As mentioned above, a fine scale simulation with this size of ε is very difficult. In order to be able to compare the solution with a reference solution, we also perform the same experiments with a larger value of ε ($\varepsilon = 10^{-3}$). This time we can generate a reference fine-scale simulation on a mesh consisting of several millions of tetrahedra and grid points. This reference solution is plotted in Figure 9 (middle picture). A numerical solution obtained with the FE-HMM for this value of ε on the same coarse mesh as before is plotted in Figure 9 (left picture). We see that we have a good qualitative agreements between both solutions and this is confirmed by numerical error estimates reported in [12]. We also see that the qualitative results for the FE-HMM are similar for the various ε (see Section 3 for a discussion on this behavior of the numerical method). Finally, we performed a simulation on a coarse grid (the same as for the FE-HMM) but with a tensor averaged in a naive way (arithmetic average). We see in Figure 9 (right picture) that the qualitative behavior is wrong and the effective conductivity obtained in this way is to large, hence we obtain an overestimation of the heat dissipation.

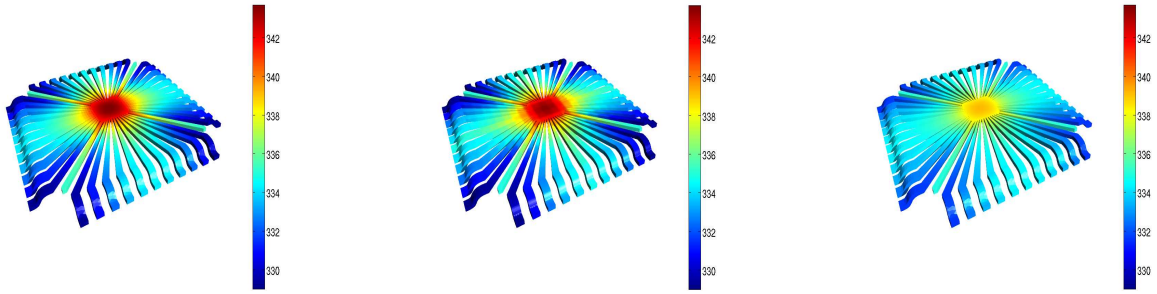


Figure 9: Temperature distribution in a leadframe. Solution of problem (12) with $\varepsilon = 10^{-3}$; FE-HMM (left picture), reference solution (middle picture), solution with averaged tensor (right solution).

2.2.4 Diffusion on rough surfaces. Diffusion on rough surfaces is a basic problem for many applications. It arises in biology as for example in the transport of lipids on the cell membrane where compartmentalization of the membrane confines the diffusion [46] or in porous media flow, where fracture of rock and pore volumes induce a local geometry which has to be taken into account for the flow transport [17]. In material science, rough surfaces arise in the study of diffusion in crystals with topological defects [21] or in the study of thermal or electrical conduction in fractures [56].

Modeling. Diffusion on rough surface can be modeled using Laplace-Beltrami like operators

$$-\Delta_{\Gamma^\varepsilon} u^\varepsilon = f \text{ in } \Gamma^\varepsilon, \quad u^\varepsilon = 0 \text{ on } \partial\Gamma^\varepsilon, \tag{14}$$

where $\Delta_{\Gamma^\varepsilon} = \nabla_{\Gamma^\varepsilon} \cdot \nabla_{\Gamma^\varepsilon}$ and $\nabla_{\Gamma^\varepsilon}$ is the tangential gradient on Γ^ε, an oscillatory surface with surface oscillations occurring at length scale ε. In some situations (as for example for crystalline objects, cell membranes, etc.), these fine structures can be obtained to high resolution by modern scanning and microscopy techniques(e.g. [40]) and the full resolution with a FEM is often out of reach.

Another source of roughness can arise from the coefficients (tensor) of the diffusion problem. Consider

$$\begin{aligned} -\nabla \cdot \left(a\left(\frac{x}{\varepsilon}, \omega\right) \nabla u^\varepsilon \right) &= f(x) \quad \text{in } \Omega, \\ u^\varepsilon(x) &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{15}$$

with random coefficients $a^\varepsilon(x) = a(x/\varepsilon, \omega) = a(T_{x/\varepsilon}\omega)$, where $\omega \in U$ and $\{T_{x/\varepsilon}\}$ is a suitable family of transformations on the sample space U (see [45, Chap. 7.1] for a precise description). Equation (15) is the typical pressure equation XS for porous media problems. In such a modeling, the natural media is seen as a statistically homogeneous realization of a random field and the permeability $a^\varepsilon(x)$ varies on an ε length scale, usually much smaller than the characteristic macroscopic length scale of observation. Again, a full resolution of the permeability field is often very costly if not infeasible.

Numerical experiments. We consider the problem (14) with a surface Γ^ε as depicted in

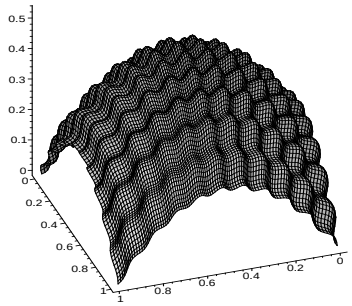


Figure 10: Example of a rough surface Γ_ε for problem (14).

Figure 10, parametrized by $F^\varepsilon(\xi) = F^0(\xi) + \varepsilon k^\varepsilon(\xi) \cdot n(\xi)$, where $n(\xi)$ is normal to the mean surface Γ^0.

Macro size	L^2 norm	H^1 norm
1/2	0.070888	0.401007
1/4	0.031271	0.174734
1/8	0.009449	0.073331
1/16	0.001673	0.025217

Table 1: Convergence of the FE-HMM for the problem on rough surface $\varepsilon = 1/50$.

While for a standard finite element method, one needs to triangulate the whole surface with a mesh which resolves the oscillation of the surface, full resolution of the fine scale in the

data is not necessary with the FE-HMM. Provided a scale separation and a self-similar fine scale distribution throughout the physical domain, the macroscopic behavior of the diffusion process can be computed on a coarse mesh (see [5] for details). For the diffusion on the surface given above, we compute a reference solution via scale resolution. The parameter $\varepsilon = 1/50$ is chosen large enough to be able to compute a reference solution with enough precision and the domain Ω is scaled to be the unit square. We emphasize again that for the FE-HMM, *any* ε can be chosen without affecting the computational cost. The FE-HMM is then applied to the problem (14) with macro meshes of 3, 5, 9, 17 points, i.e., macro meshsizes of $1/2, 1/4, 1/8, 1/16$. We compare in Table 1 the L^2 projection of the reference solution with the FE-HMM solution in the L^2 and H^1 norms. We see that we can capture the right macro diffusion process with substantially fewer degrees of freedom than needed with a standard FEM.

We next consider the problem (15) with random coefficients. We chose a^ε to be a log-normal stochastic field with mean-zero, variance $\sigma = 1$ and correlation length $\varepsilon_1 = 0.02, \varepsilon_2 = 0.03$. We generate a realization of this stochastic field by the moving ellipse averaging method

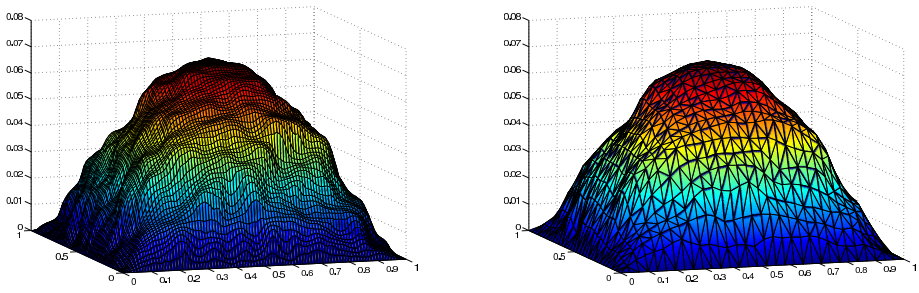


Figure 11: Comparison of the fine scale solution (pressure profile) of problem (15) with random coefficients on a 1024×1024 grid and the FE-HMM solution on a 32×32 grid.

(see [64, Section 4.1] for a description). We then compute a reference solution on a fine 1024×1024 grid and compare the solution with the FE-HMM (the reference solution can be seen as finite difference version of (15) [10]; the leading order behavior is determined by the homogenized solution that we aim at capturing). We compute a solution with the FE-HMM on a coarse 32×32 grid, choosing a sampling domain of size 0.06×0.06. It can be seen in Figure 11 that the solution obtained from the FE-HMM on the aforementioned coarse grid is in good qualitative agreement with the solution of the standard FEM on the fine grid (1024×1024 grid).

3. The Finite Element Heterogeneous Multiscale Method (FE-HMM)

In this section we discuss in details the FE-HMM. This method is based on the framework introduced in [32]. In the context of PDEs, the first numerical method based on HMM, the so-caled FD-HMM has been obtained in [2], where a finite difference (FD) method has been derived and analyzed for parabolic problems. The FE-HMM was first discussed in [5] and [33] for (non-uniformly) periodic problems. In [33] nonlinear and stochastic problems were discussed and partially analyzed. In [5] robust convergence rates (i.e. independent of the small scale ε) were obtained for linear problems. Both the analysis in [5] and [33]

were obtained for a semi-discrete numerical method, i.e., assuming that the small scale were solved exactly. Such assumptions were commonly made in the analysis of most of the existing multiscale methods for PDEs [42],[35],[52]. We note that in [53] macro and micro error were first separated and quantitatively estimated, although not for the HMM and restricted to elliptic problems with uniformly periodic tensor and unbounded domains (this analysis cannot be easily generalized to other multiscale scenarios).

The first fully discrete analysis for HMM was obtained in [6], where the error propagation across scales has been analyzed and optimal a-priori bounds have been obtained. This analysis has later been extended to elasticity problems [8] and to advection-diffusion problems [7]. The importance of a fully discrete analysis became also clear for other type of multiscale methods for which such analysis have later been proposed [15], [44]. The clear separation of micro and macro errors for the HMM not only led to a better understanding of the complexity of the numerical method, but also paved the way for a "goal oriented" coupling for HMM, i.e. the *coupling* of different type of solvers at different scales. Such hybrid couplings have been investigated in [10], where a multiscale method based on a FEM (macro scale) and a spectral method (micro scale) has been proposed and analyzed (the FES-HMM), and in [11], where a multiscale method based on a discontinuous Galerkin finite element method (macro-scale) and a FEM (micro scale) has been proposed and analyzed (the DG-HMM).

We start this section by briefly discussing homogenization problems (Section 3.1), we then introduce the FE-HMM (Section 3.2) and discuss the analysis of the method (Section 3.3).

3.1 Homogenization problems.
We recall here briefly the class of problems for which we want to propose a multiscale algorithm. Examples for such problems have already been discussed in Section 2. For simplicity and clarity of the presentation, we restrict ourself to multiscale elliptic problems. Similar ideas as developed here apply to problem in elasticity or parabolic problems as mentioned before.

We consider a convex polygonal domain $\Omega \in \mathbb{R}^d,\ d = 1, 2, 3$ with a boundary $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N$ where Dirichlet conditions are imposed on $\partial\Omega_D$ and Neumann conditions on $\partial\Omega_N$. For simplicity we assume that $\partial\Omega_D \cap \partial\Omega_N = \emptyset$ and that $\partial\Omega_D$ has positive measure, but pure Neumann or mixed boundary conditions could be considered without further difficulties. Given $f \in L^2(\Omega),\ g_D \in H^1(\Omega),\ g_N \in L^2(\partial\Omega_N)$, we consider the second-order elliptic equation

$$\begin{aligned} -\nabla \cdot (a^\varepsilon \nabla u^\varepsilon) &= f \quad \text{in } \Omega, \\ u^\varepsilon &= g_D \quad \text{on } \partial\Omega_D, \\ n \cdot (a^\varepsilon \nabla u^\varepsilon) &= g_N \quad \text{on } \partial\Omega_N, \end{aligned} \tag{16}$$

where a^ε is symmetric, satisfies $a^\varepsilon(x) \in (L^\infty(\Omega))^{d\times d}$ and is uniformly elliptic and bounded, i.e.,

$$\exists \lambda, \Lambda > 0 \text{ such that } \lambda|\xi|^2 \le a^\varepsilon(x)\xi \cdot \xi \le \Lambda|\xi|^2,\ \forall \xi \in \mathbb{R}^d,\ \forall \varepsilon, \tag{17}$$

where ε represents a small scale in the problem that characterizes the multiscale nature of the tensor $a^\varepsilon(x)$. An application of Lax-Milgram theorem gives us a family of solution which is bounded in $H^1_0(\Omega)$ independently of ε. The variational problem corresponding to (16) is

the following: find $u^\varepsilon \in H^1_D(\Omega)$ such that

$$B^\varepsilon(u^\varepsilon, v) := \int_\Omega a^\varepsilon \cdot \nabla u^\varepsilon \nabla v dx = \int_\Omega f v dx + \int_{\partial\Omega_N} g_N v dx - \int_\Omega a^\varepsilon \cdot \nabla g_D \nabla v dx =: F(v), \quad (18)$$

for all $v \in H^1_D(\Omega)$, where $H^1_D(\Omega) := \{v \in H^1(\Omega); v = 0 \text{ on } \partial\Omega_D\}$. A finite element discretization of this variational problem is standard and is briefly described below. Let $\mathcal{T}_h$ be a partition of Ω in simplicial or quadrilateral elements K of diameter h_K and denote $h = \max_{K\in\mathcal{T}_h} h_K$. In this section we will always assume that the triangulation is admissible and shape regular, i.e.,

- $\bigcup_{K\in\mathcal{T}_h} K = \bar{\Omega}$, the intersection of two elements is either empty, exactly one vertex or a common face (admissible),
- $\exists \kappa > 0$ such that $h_K/\rho_K \le \kappa$, where ρ_K is the diameter of the largest circle contained in K (shape regular).

The first condition can be relaxed for other types of discretizations as we will see in Section 4. For a partition as described above, we define a finite dimensional subspace of $H^1_D(\Omega)$ by

$$V^l_D(\Omega, \mathcal{T}_h) = \{v^h \in H^1_D(\Omega);\ v^h|_K \in \mathcal{R}^l(K),\ \forall K \in \mathcal{T}_h\}, \quad (19)$$

where $\mathcal{R}^l(K)$ is the space $\mathcal{P}^l(K)$ of polynomials on K of total degree at most l if K is a simplicial FE, or the space $\mathcal{Q}^l(K)$ of polynomials on K of degree at most l in each variables if K is a rectangular FE. We assume that the partition $\mathcal{T}_h$ is regular (see [25] for details). The solution of the discretized problem reads: find $u^h \in V^l_D(\Omega, \mathcal{T}_h)$ such that

$$B^\varepsilon(u^h, v^h) = F^\varepsilon(v^h) \quad \forall v^h \in V^l_D(\Omega, \mathcal{T}_h). \quad (20)$$

Although standard, there is a major issue with this approach: solving (18) with a standard FEM needs usually to resolve the smallest scale present in the problem (denoted here by ε). Roughly speaking, assuming that the smoothness of the data and the domain are such that $u \in H^{l+1}(\Omega)$, then the a-priori estimate $\|u\|_{H^{l+1}(\Omega)} \le C\varepsilon^{-l}\|f\|_{H^{l-1}}$ holds, where C is independent of ε. Then, the sharp a-priori error bound between the solution u^ε of (18) and the FE solution u^h of (20)

$$\|u^\varepsilon - u^h\|_{H^1(\Omega)} \le C\left(\frac{h}{\varepsilon}\right)^l \|f\|_{H^{l-1}(\Omega)},$$

can be derived following classical results [25]. This means that the meshsize should satisfy $h < \varepsilon$. Thus, if ε is small, the cost associated with the FEM (20) will be prohibitive.

Homogenization method. As mentioned in the introduction, an effective dynamics for a multiscale PDE can be described by using homogenization theory. Homogenization theory has been an active field of research for the past 30 years. Among the huge literature we mention three books [22],[26],[45] where the interested reader can find more material on the subject including details of the brief discussion which follows.

Without further assumptions on the heterogeneities of the tensor $a^\varepsilon(x)$ the theory of $G-$convergence introduced by De Giorgi and Spagnolo [38] [3] can be used to show that there

[3] In the general non symmetric case one can use the theory of $H-$convergence introduced by Tartar in 1977 and developed by Murat and Tartar [57].

exists a symmetric tensor $a^0(x)$ and a subsequence of $\{u^\varepsilon\}$ which weakly converges to an element $u^0 \in H_0^1(\Omega)$ solution of the so-called homogenized or upscaled problem

$$\begin{aligned} -\nabla \cdot \left(a^0 \nabla u^0\right) &= f \quad \text{in } \Omega, \\ u^0 &= g_D \quad \text{on } \partial\Omega_D, \\ n \cdot \left(a^0 \nabla u^0\right) &= g_N \quad \text{on } \partial\Omega_N, \end{aligned} \tag{21}$$

where the homogenized tensor $a^0(x)$ again satisfies $\lambda|\xi|^2 \leq a^0(x)\xi \cdot \xi \leq \Lambda|\xi|^2,\ \forall \xi \in \mathbb{R}^d$. Under additional assumptions on the small scale such as periodicity[4], explicit equations are available to compute the homogenized tensor given by

$$a^0_{ij}(x) = \int_Y \left(a_{ij}(x,y) + \sum_{k=1}^{d} a_{ik}(x,y) \frac{\partial \chi^j}{\partial y_k}(x,y) \right) dy. \tag{22}$$

Here $\chi^j(x,\cdot),\ j = 1, \ldots, d$ are defined to be the unique solutions of the cell problems

$$\int_Y \nabla \chi^j(x,y) a(x,y) \nabla v(y) dy = -\int_Y (a(x,y)e_j)^T \nabla v(y) dy, \quad \forall v(y) \in W^1_{per}(Y), \tag{23}$$

where $(e_j)_{j=1}^d$ is the canonical basis of $\mathbb{R}^d$. Notice that a Poincaré-Wirtinger inequality is available in $W^1_{per}(Y)$ (see [26, Chap. 3]), hence the existence and uniqueness of the problem (23) is guaranteed by the Lax-Milgram theorem. Strong error estimates between the solutions of (16) and (21) are available in the L^2 norm [45, Sect. 1.4]

$$\|u^\varepsilon - u^0\|_{L^2(\Omega)} \leq C\varepsilon. \tag{24}$$

Due to the ε oscillations of the fine scale solution, strong error in the H^1 norm can usually not be obtained since the gradients of the oscillations are in general not $\mathcal{O}(\varepsilon)$ quantities. The homogenized solution needs thus to be "corrected" through information of the fine scale. This can be done by defining a corrector given by

$$u_1(x, x/\varepsilon) = \sum_{j=1}^{n} \chi^j(x, x/\varepsilon) \frac{\partial u^0(x)}{\partial x_j}, \tag{25}$$

where the functions $\chi^j(x, x/\varepsilon)$ are given by (23) and we then have [45, Sect. 1.4]

$$\|u^\varepsilon - (u^0 + \varepsilon u_1(x, x/\varepsilon))\|_{H^1(\Omega)} \leq C\sqrt{\varepsilon}, \tag{26}$$

where a boundary layer term is responsible for the $\sqrt{\varepsilon}$ (instead of ε) convergence rate (notice that u_1 does not satisfy the boundary conditions of the problem (16)). Some regularity on u^0 and $\chi^j(x,\cdot)$ is needed for the estimates (24) and (26) and we refer to [45, Chap.1.4] (see also the discussion in [42, Remark 3.3] and [44, Section 3.4]).

3.2 FE-HMM: the numerical algorithm

The so-called finite element heterogeneous multiscale method (FE-HMM) aims at capturing

[4]For example $a^\varepsilon = a(x, x/\varepsilon) = a(x,y)$ is Y-periodic in y, where Y is for example the unit cube $Y = (0,1)^d$, and $a(x,y) \in C[\Omega; L^\infty_{per}(Y)]$.

the homogenized (coarse) solution u^0 of (21) without knowing or precomputing $a^0(x)$. For simplicity of notation, we suppose here that $\partial\Omega_N = \emptyset,\ g_D = 0$, but we emphasize that the FE-HMM is not restricted to this special case as already seen in the examples of Section 2. We describe here the main components of the FE-HMM: the macro and micro finite element (FE) spaces and the modified macro bilinear form based on quadrature formula (QF).

Macro finite element space. We consider

$$V^p(\Omega, \mathcal{T}_H) = \{v^H \in H_0^1(\Omega);\ u^H|_K \in \mathcal{R}^p(K),\ \forall K \in \mathcal{T}_H\}, \tag{27}$$

a finite element (FE) space similar to (19) but defined on macro elements K with size H allowed here to be much larger than ε. Within each macro element $K \in \mathcal{T}_H$ we consider, for $j = 1, \ldots, J$,

- integration nodes $x_{j,K} \in K$,
- sampling domains $K_\delta(x_{j,K}) = x_{j,K} + \delta I$, where $I = (-1/2, 1/2)^d$ and $\delta \geq \varepsilon$,
- quadrature weights $\omega_{j,K}$.

Quadrature formula. Let $\hat{K}$ be the reference element and consider for any element of the triangulation the mapping F_K (a C^1-diffeomorphism) such that $K = F_K(\hat{K})$. The set $\{\hat{x}_j, \hat{\omega}_j\}_{j=1}^J$ is a quadrature formula on $\hat{K}$ chosen such that

$$\int_{\hat{K}} \hat{p}(\hat{x})d\hat{x} = \sum_{j\in J} \hat{\omega}_j \hat{p}(\hat{x}_j)\ \forall \hat{q}(\hat{x}) \in \mathcal{R}^\sigma(\hat{K}), \tag{28}$$

where we will assume that the weights satisfy $\hat{\omega}_j > 0$. The QF (28) induces a QF over K via $x_{j,K} = F_K(\hat{x}_j),\ \omega_{j,K} = \hat{\omega}_j \det(\partial F_K),\ j = 1 \ldots, J$. The conditions on the QF to ensure that a FEM with numerical quadrature [5] converges to the exact solution with the same rate than a FEM with exact integration, have been studied by Ciarlet and Raviart (see [25, Chap. 4.1]). We briefly recall these conditions as they will be important for the FE-HMM.

Ellipticity condition. If a QF is used to compute a bilinear form (see for example (20)), then the ellipticity is no longer guaranteed unless suitable conditions on the QF are satisfied. Given a polynomial space $\mathcal{R}^\sigma(\hat{K})$ (either $\mathcal{P}^\sigma(\hat{K})$ or $\mathcal{Q}^\sigma(\hat{K})$ polynomials as defined in (19)) we will require that

$$\sqrt{\sum_{j\in J} \omega_j |\nabla \hat{p}(\hat{x}_j)|^2} \quad \text{is a norm on the finite dimensional space } \mathcal{R}^\sigma(\hat{K})/\mathcal{R}^0(\hat{K}). \tag{29}$$

The above property holds if the nodes $\{\hat{x}_j\}_{j=1}^J$ contain a so-called unisolvent set for the derivatives of the considered polynomial set (see [25, Thm. 4.1.2 and Ex. 4.1.7]) which means

$$\forall \hat{p} \in \widetilde{\mathcal{P}}(\hat{K}),\ \forall i = 1, \ldots, d \quad \text{if} \quad \frac{\partial \hat{p}}{\partial \hat{x}_i}(\hat{x}_j) = 0,\ j = 1, \ldots, L \quad \text{then} \quad \frac{\partial \hat{p}}{\partial \hat{x}_i} \equiv 0,$$

[5] i.e. when the integral in (20) are replaced by the above QF

where $\widetilde{\mathcal{P}}(\hat{K}) = \mathcal{P}^{p-1}(\hat{K})$ for simplicial FEs, while $\widetilde{\mathcal{P}}(\hat{K}) = \mathcal{Q}^p(\hat{K}) \cap \mathcal{P}^{dp-1}(\hat{K})$ for rectangular FEs.

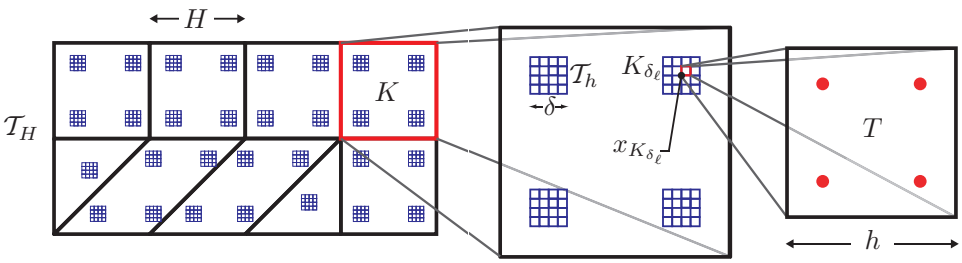
Approximation condition. Let $u_{h,QF}$ be the FE solution of a variational elliptic problem where all the integrals arising in the problem are computed with a QF. We will require that the QF is chosen such that the standard error estimates for a FEM hold. Assuming sufficient regularity of the solution this reads

$$\|u - u_{h,QF}\|_{H^1(\Omega)} \leq Ch^p, \quad \|u - u_{h,QF}\|_{L^2(\Omega)} \leq Ch^{p+1}, \tag{30}$$

where the approximation $u_{h,QF}$ is continuous and piecewise in $\mathcal{P}^p(K)$ or $\mathcal{Q}^p(K)$. For $p > 1$ the estimates (30) hold if the QF is exact for $\mathcal{P}^{2p-2}(\hat{K})$ (simplicial FE) or if the QF is exact for $\mathcal{Q}^{2p-1}(\hat{K})$ (quadrilateral FE). The same conditions apply if $p = 1$ for the estimate in the H^1 norm while for the estimate in the L^2 norm, the QF should be exact for $\mathcal{P}^1(\hat{K})$ (simplicial FE) or for $\mathcal{Q}^2(\hat{K})$ (quadrilateral FE). We refer to [24, Thms. 9 and 11] and [25, Chap. 4.1] for details.

Example. For piecewise linear elements $J = 1$, $\omega_K = |K|$ and x_K is chosen to be located at the barycenter of the simplicial K. For bilinear elements, $J = 4$ and $\{\omega_{j,K}, x_{j,K}\}_{j=1}^4$ is the two points Gauss quadrature rule given by $\omega_{j,K} = |K|/4, x_{j,K} = F_K(1/2 \pm \sqrt{3}/6, 1/2 \pm \sqrt{3}/6)$, where F_K is the affine mapping such that $F_K(\hat{K}) = K$ and $\hat{K} = (0,1)^d$ (see Figure 12).

Macro bilinear form. For a discretization in the coarse FE space (27) we need to modify



(a) macro-problem (b) micro-problem (c) quadrature for the microproblem

Figure 12: Example of a macro FE space made of triangles and quadrilaterals with sampling domains at integration nodes. A micro FEM is defined at each sampling domain as depicted in (b). Numerical integration must usually also be performed on the micro FEM as depicted in (c).

the bilinear form (20). For $v^H, w^H \in V^p(\Omega, \mathcal{T}_H)$ we define

$$B_H(v^H, w^H) = \sum_{K \in \mathcal{T}_H} \sum_{j=1}^{J} \frac{\omega_{j,K}}{|K_\delta(x_{j,K})|} \int_{K_\delta(x_{j,K})} a^\varepsilon(x) \nabla v_j^h \cdot \nabla w_j^h dx, \tag{31}$$

where v_j^h, w_j^h are micro functions defined on sampling domains $K_\delta(x_{j,K})$ by the problem (35) below and the factor $|K_\delta(x_{j,K})|$ (the measure of $K_\delta(x_{j,K})$) gives the appropriate weight for the contribution of the integrals defined on $K_\delta(x_{j,K})$ instead of K.

Micro FE space. We consider a (micro) partition $\mathcal{T}_h$ of each sampling domain $K_\delta(x_{j,K})$ in simplicial or quadrilateral elements Q of diameter h_Q and denote $h = \max_{i \in \mathcal{T}_h} h_Q$. For this

partition we define a micro FE space

$$S^q(K_\delta, \mathcal{T}_h) = \{z_h \in W(K_\delta(x_{j,K}));\ z_h|_T, \in \mathcal{R}^r(Q),\ Q \in \mathcal{T}_h\}, \tag{32}$$

where $W(K_\delta(x_{j,K}))$ is a Sobolev space whose choice sets the boundary conditions for the micro problems and thus determines the type of coupling between micro and macro problems. Several choices are possible for the coupling condition. We consider the following cases

$$W(K_\delta(x_{j,K})) = W^1_{per}(K_\delta(x_{j,K}) \text{ and we set } S^l_P(K_\delta, \mathcal{T}_h) := S^q(K_\delta, \mathcal{T}_h), \tag{33}$$
$$W(K_\delta(x_{j,K})) = H^1_0(K_\delta(x_{j,K}) \text{ and we set } S^l_D(K_\delta, \mathcal{T}_h) := S^q(K_\delta, \mathcal{T}_h). \tag{34}$$

Micro problem. Find for every macro element K the additive contribution to the macro stiffness matrix by computing the micro functions v^h_j (respectively w^h_j) on the sampling domain $K_\delta(x_{j,K}),\ j = 1, \ldots, J$ such that $(v^h_j - v^H_{lin,j}) \in S^q(K_\delta(x_{j,K}), \mathcal{T}_h)$ and

$$\int_{K_\delta(x_{j,K})} a^\varepsilon(x)\nabla v^h_j \cdot \nabla z^h dx = 0 \quad \forall z^h \in S^q(K_\delta(x_{j,K}), \mathcal{T}_h), \tag{35}$$

where

$$v^H_{lin,j}(x) = v(x_{j,K})^H + (x - x_{j,K}) \cdot \nabla v^H(x_{j,K}), \tag{36}$$

is a linearization of the macro function v^H at the integration node $x_{j,K}$ (of course for piecewise linear functions $v^H_{lin,(x_{j,K})} = v^H$). Notice that there is a slight abuse of notation in the above definition and we should use $v^H_{lin,x_{j,K}}(x)$ instead of $v^H_{lin,j}(x)$, but we will avoid carrying this heavy notation when no confusion can occur.

Remark 1 *For a tensor $a^\varepsilon(x) = a(x, x/\varepsilon)$ with explicit scale separation, it is preferable to collocate the slow variable at the integration points $a(x_{j,K}, x/\varepsilon)$ in both the macro and micro bilinear forms (31) and (35). In the periodic case, choosing δ as an integer multiple of ε gives robust, i.e. independent of ε, convergence results (see [5, App. A]).*

Remark 2 *Of course numerical quadrature must also be used in general for the micro problem (35), but there, standard QF can be used and usual error estimates apply (see [25, Chap. 4.1]). The reason why we insist on QF for the macro scheme (31), is that this modified bilinear form itself, and in turn the HMM strategy, rely on QF defined on "sampling domains".*

The Multiscale Method. In view of the modified bilinear form defined in (31), the FE-HMM reads: find $u_H \in V^p(\Omega, \mathcal{T}_H)$ such that

$$B_H(u^H, v^H) = F(v^H) \quad \forall v^H \in V^p(\Omega, \mathcal{T}_H). \tag{37}$$

Several remarks are in order. First, the computational saving compared to solving (18) is clear since instead of solving the fine scale on the whole computational domain (as required for (18) with $h < \varepsilon$), in the FE-HMM, we only solve the fine scale on sampling domains K_δ, usually of much smaller size than the macro meshsize H. Second, the coupling between micro and macro methods allows for much flexibility in the choice of the macro and micro discretization spaces in order to balance the computational cost between micro and macro

solver [6] or to obtain specific qualitative properties of the numerical solution at a given scale.

Other type of macro-micro solvers for HMM

The FE-HMM is not restricted to continuous FE discretization for the macro and micro solver as described above. The HMM strategy offers much flexibility in the choice of the type of solvers used at a given scale. Several methods have been proposed in this direction. We mention [10], where pseudo-spectral methods have been used for the micro problems allowing (provided enough regularity) for spectral or exponential convergence in the micro methods and [11], where a discontinuous Galerkin FEM has been used for the macro space, allowing for nonmatching meshes, approximation flexibility and mass and flux conservation. We will comment in Section 4 on these developments.

Post-processing procedure

The primary goal of the FE-HMM is to capture the effective solution u_0 of (21). While $u^H \to u^\varepsilon$ in the L^2 norm, such a convergence cannot be obtained in the energy norm. An energy approximation can nevertheless be constructed by using a post-processing procedure and extending periodically on the whole element K the function $(u^h - u^H_{lin})$ capturing the micro oscillations and available in $K_\delta(x_{j,K}) \subset K$. This will be discussed at the end of Section 3.3.

3.3 FE-HMM: fully discrete a priori error analysis

In this section we give a detailed analysis of the FE-HMM method. First, we show that the bilinear form (31) is coercive. This implies the existence and uniqueness of a solution of the problem (37) and can be done without specific assumptions on a^ε (of course we assume (17)). Second, we derive a priori estimates by decomposing the error as

$$\|u^0 - u^H\| \leq e_{MAC} + e_{MOD} + e_{MIC}, \tag{38}$$

where $e_{MAC}, e_{MOD}, e_{MIC}$ denote the macro, modeling and micro errors and $\|\ \|$ denotes the H^1 or L^2 norm. To estimate e_{MOD} and e_{MIC} some knowledge of the homogenized problem is needed and we will assume (non-uniform) periodicity of the tensor a^ε. We emphasize that the numerical algorithm, i.e. the FE-HMM itself, is not restricted to such assumptions and can be applied to more general problems (however scale separation and self-similarity are needed for the strategy to make sense). The careful analysis of the fully discrete numerical scheme besides giving precise convergence rate in the periodic case also give some indication of the behavior of the method in the more general non-periodic setting. Indeed, the various components of the error, the influence of the boundary conditions in the coupling of macro and micro methods are likely to be present also for more general problems. This analysis is thus a fundamental step towards designing robust and reliable numerical methods based on macro and micro solvers. The analysis presented in this section is based on [33],[2],[5](for e_{MAC} and e_{MOD}) and [6],[7],[9] for (for e_{MIC}).

Here and in what follows we will simplify the notation for the sampling domain and use K_δ or K_{δ_j} instead of $K_\delta(x_{j,K})$ when no confusion can occur. Notice also that the micro FE space $S^q(K_\delta, \mathcal{T}_h)$ (see (32)) denotes either $S^q_D(K_\delta(x_{j,K}), \mathcal{T}_h)$ or $S^q_P(K_\delta(x_{j,K}), \mathcal{T}_h)$ when proving results which hold for both type of coupling.

Assumptions. As mentioned is Section 3.2, the macro bilinear form in the FE-HMM is

[6] Remember that the method also depends on a micro mesh, thus $H \to 0$ and $h \to 0$ are necessary for convergence.

based on QF. In what follows, we will always assume that the QF upon which the bilinear form (31) is constructed satisfies (29) and (30), and these assumptions will be implicitly assumed in the various results below when needed.

3.3.1 Coercivity and well-posedness. We first notice that the micro problem (35) has a unique solution. This follows from (17), the Poincaré or the Poincaré-Wirtinger inequality for $S^q_D(K_\delta(x_{j,K}),\mathcal{T}_h)$ and $S^q_P(K_\delta(x_{j,K}),\mathcal{T}_h)$, respectively, and the Lax-Milgram Lemma. It follows that the form (31) is indeed a bilinear form on $V^p(\Omega,\mathcal{T}_H)$.

The constrained micro calculation in sampling domains (35) sets a coupling between micro and macro functions. The following lemma gives an energy equivalence between these functions on sampling domains and is the basis for proving the coercivity of (31).

Lemma 3 *Let v^h be the solution of (35) constrained by v^H_{lin} the linearization of $v^H \in V^p(\Omega,\mathcal{T}_H)$ defined in (36). Then,*

$$\|\nabla v^H_{lin}\|_{L^2(K_\delta)} \le \|\nabla v^h\|_{L^2(K_\delta)} \le \sqrt{\frac{\Lambda}{\lambda}}\|\nabla v^H_{lin}\|_{L^2(K_\delta)}, \tag{39}$$

where λ,Λ are defined in (17).

Proof. A direct calculation gives

$$\begin{aligned}\int_{K_\delta}(\nabla v^h-\nabla v^H_{lin})\cdot(\nabla v^h-\nabla v^H_{lin})dx &= \int_{K_\delta}|\nabla v^h|^2dx+\int_{K_\delta}|\nabla v^H_{lin}|^2dx\\ &- 2\int_{K_\delta}\nabla v^H_{lin}\cdot\nabla v^h dx.\end{aligned}$$

By noting that

$$\int_{K_\varepsilon}\nabla v^H_{lin}\cdot\nabla v^h dx=\nabla v^H_{lin}\cdot\int_{K_\delta}(\nabla v^h-\nabla v^H_{lin})dx+\int_{K_\delta}|\nabla v^H_{lin}|^2dx=\int_{K_\delta}|\nabla v^H_{lin}|^2dx,$$

where we used that ∇v^H_{lin} is constant and that $(v^H_{lin}-v^h)|_{\partial K_\delta}$ vanishes for periodic or Dirichlet coupling (see (32)), we obtain the left inequality of (39). For the second inequality, we observe that

$$\begin{aligned}\int_{K_\delta}a^\varepsilon(x)\nabla v^h\cdot\nabla v^h dx &= \int_{K_\delta}a^\varepsilon(x)\nabla v^H_{lin}\cdot\nabla v^H_{lin}dx\\ &- \int_{K_\delta}a^\varepsilon(x)\,(\nabla v^h-\nabla v^H_{lin})\cdot(\nabla v^h-\nabla v^H_{lin})dx,\end{aligned}$$

where we used repeatedly that $\int_{K_\delta(x_{j,K})}a^\varepsilon(x)\nabla v^h\cdot(\nabla v^h-\nabla v^H_{lin})dx=0$ since v^h is a solution of (35) and $(\nabla v^h-\nabla v^H_{lin})\in S^q(K_\delta,\mathcal{T}_h)$. Using the ellipticity assumption (17) gives the result. □

Remark 4 *The assertion of the above lemma remains true if (35) is solved exactly, i.e. in* $W(K_\delta(x_{j,K}))$ *instead of* $S^q(K_\delta, \mathcal{T}_h)$*. Let* v *be this solution constrained by* v^H_{lin} *(as in the above lemma). Then,*

$$\|\nabla v^H_{lin}\|_{L^2(K_\delta)} \leq \|\nabla v\|_{L^2(K_\delta)} \leq \sqrt{\frac{\Lambda}{\lambda}}\|\nabla v^H_{lin}\|_{L^2(K_\delta)}. \tag{40}$$

With the help of Lemma 3, we can prove that the bilinear form (31) upon which the FE-HMM is defined is uniformly elliptic and bounded.

Lemma 5 *The bilinear form (31) satisfies*

$$B_H(v^H, v^H) \geq C\|v^H\|^2_{H^1(\Omega)},\ |B_H(v^H, w^H)| \leq C\|v^H\|_{H^1(\Omega)}\|w^H\|_{H^1(\Omega)}, \tag{41}$$

for all $v^H, w^H \in V^p(\Omega, \mathcal{T}_H)$*, where the constant* C *only depends on the QF (see (29)), (17) and the domain* Ω*.*

Proof. Let $v^H \in V^p(\Omega, \mathcal{T}_H)$ and using the notation K_{δ_j} for $K_\delta(x_{j,K})$, we have

$$\begin{aligned} B_H(v^H, v^H) &= \sum_{K\in\mathcal{T}_H}\sum_{j=1}^{J}\frac{\omega_{j,K}}{|K_{\delta_j}|}\int_{K_{\delta_j}} a^\varepsilon(x)\nabla v^h_j \cdot \nabla v^h_j dx \\ &\geq \lambda\sum_{K\in\mathcal{T}_H}\sum_{j=1}^{J}\omega_{j,K}|\nabla v^H_{lin}|^2 \geq C\sum_{K\in\mathcal{T}_H}\|\nabla v^H\|^2_{L^2(K)} \geq C\|v^H\|^2_{H^1(\Omega)}, \end{aligned}$$

where we have used the ellipticity (17) of the tensor and Lemma 3 to obtain the first inequality and the identity $\nabla v^H_{lin}(x) \equiv \nabla v^H(x_{j,K})$ for $x \in K$, the assumption on the quadrature formula and the Poincaré inequality for the second inequality. It is clear that (31) is bounded on $V^p(\Omega, \mathcal{T}_H)$ since it is a finite dimensional space. To show that the bound is uniform in ε, we use Lemma 3, the bound (17) and the fact that the right hand side of (29) defines a norm on the finite dimensional polynomial quotient space. □

In view of Lemma 5 and the Lax-Milgram lemma we obtain the existence and uniqueness of the problem (37).

Theorem 6 *The problem (37) has a unique solution which satisfies*

$$\|u^H\|_{H^1(\Omega)} \leq C\|f\|_{L^2(\Omega)}. \tag{42}$$

3.3.2 A-priori estimates. In view of (38), we have to estimate the macro, micro and modeling errors.

Macro error. Besides (17), we do not need any other assumptions on the fine scale tensor a^ε. Notice that in the framework of G or H convergence, the ellipticity and boundedness of $a^0(x)$ is guaranteed, but explicit expressions for this tensor as (22) are in general not available [26, Chap. 13]. Define a bilinear form on $V^p(\Omega, \mathcal{T}_H) \times V^p(\Omega, \mathcal{T}_H)$ for the problem (21), using the QF (28):

$$B_{0,H}(v^H, w^H) = \sum_{K\in\mathcal{T}_H}\sum_{j=1}^{J}\omega_{j,K}\ a^0(x_{j,K})\nabla v^H(x_{j,K}) \cdot \nabla w^H(x_{j,K})dx. \tag{43}$$

Define $u^{0,H}$ to be the solution of

$$B_{0,H}(u^{0,H}, v^H) = F(v^H) \quad \forall v^H \in V^p(\Omega, \mathcal{T}_H). \tag{44}$$

The next result follows from (30).

Proposition 7 *Suppose that the bilinear form (43) is based on the QF (28) and that (29) and (30) hold. Suppose further that the solution of the problem (21) satisfies $u^0 \in H^{p+1}(\Omega)$. Then,*

$$e_{MAC,H^1} := \|u^0 - u^{0,H}\|_{H^1(\Omega)} \leq CH^p, \quad e_{MAC,L^2} := \|u^0 - u^{0,H}\|_{L^2(\Omega)} \leq CH^{p+1}. \tag{45}$$

Micro error. For this part of the error, we are concerned with the propagation of the discretization error of the micro problem (35) at the macro scale. Here again, as for the estimation of the macro error no assumptions on the fine scale tensor a^ε are needed besides (17). We define a bilinear form on $V^p(\Omega, \mathcal{T}_H) \times V^p(\Omega, \mathcal{T}_H)$ by

$$\bar{B}_H(v^H, w^H) = \sum_{K \in \mathcal{T}_H} \sum_{j=1}^{J} \frac{\omega_{j,K}}{|K_\delta(x_{j,K})|} \int_{K_\delta(x_{j,K})} a^\varepsilon(x) \nabla v_j \cdot \nabla w_j dx, \tag{46}$$

where v_j, w_j are the solutions of (35) in the 'exact" Sobolev space $W(K_\delta(x_{j,K}))$ instead the FE space $S^q(K_\delta, \mathcal{T}_h)$. Define $\bar{u}^H$ to be the solution of

$$\bar{B}_H(\bar{u}^H, v^H) = F(v^H) \quad \forall v^H \in V^p(\Omega, \mathcal{T}_H). \tag{47}$$

The next result gives an estimation of the micro error.

Proposition 8 *Let $u^H, \bar{u}^H$ be the solutions of (37) and (47), respectively, with the same coupling condition, either (33) or (34). Suppose that the assumptions of Lemma 10 (see below) hold. Then*

$$e_{MIC,H^1} := \|u^H - \bar{u}^H\|_{H^1(\Omega)} \leq C \left(\frac{h}{\varepsilon}\right)^{2q}. \tag{48}$$

Proof. Denoting by $w^H = u^H - \bar{u}^H$, and using (41) we have

$$C\|u^H - \bar{u}^H\|^2_{H^1(\Omega)} \leq B_H(u^H - \bar{u}^H, w^H) = \bar{B}_H(\bar{u}^H, w^H) - B_H(\bar{u}^H, w^H), \tag{49}$$

and thus

$$\|u^H - \bar{u}^H\|_{H^1(\Omega)} \leq C \sup_{w^H \in V^p(\Omega, \mathcal{T}_H)} \frac{|\bar{B}_H(\bar{u}^H, w^H) - B_H(\bar{u}^H, w^H)|}{\|w^H\|_{H^1(\Omega)}}. \tag{50}$$

Using Lemma 10 given below proves the result. □

To estimate the difference between the bilinear forms B_H and $\bar{B}_H$, we have to study the approximation error in the micro problem (35). We need first some preparation. Define $\eta^{i,h}(x),\ i=1,\ldots,d$ to be the solution of

$$\int_{K_{\delta_j}} a^\varepsilon(x)\nabla\eta^{i,h}\cdot\nabla z^h dx = -\int_{K_{\delta_j}} a^\varepsilon(x)e_i\cdot\nabla z^h dx \quad \forall z^h\in S^q(K_{\delta_j}),\mathcal{T}_h), \tag{51}$$

where $(e_i)_{i=1}^d$ is the canonical basis of $\mathbb{R}^d$ (notice that we used the notation K_{δ_j} instead of $K_\delta(x_{j,K})$). Likewise, let $\eta^i(x)\ i=1,\ldots,d$ be the (non-discretized) solution of (51) in $W(K_{\delta_j})$ instead of $S^q(K_{\delta_j},\mathcal{T}_h)$. Then, the solution v_j^h of (35) and the solution v_j of (35) (in $W(K_{\delta_j})$) can be written as

$$v_j^h(x)=v_{lin,j}^H(x)+\sum_{i=1}^d \eta^{i,h}(x)\frac{\partial v_{lin,j}^H(x)}{\partial x_i},\quad v_j(x)=v_{lin,j}^H(x)+\sum_{i=1}^d \eta^{i}(x)\frac{\partial v_{lin,j}^H(x)}{\partial x_i}, \tag{52}$$

respectively. This can easily be seen just by replacing the above expressions in (51) and using the uniqueness its solutions.

Remark 9 *For the case of a (non-uniformly) periodic tensor $a^\varepsilon(x)=a(x,x/\varepsilon)$, the function $v_j(x)$ in (52) can be written as*

$$v_j(x)=v_{lin,j}^H(x)+\sum_{i=1}^d \varepsilon\chi^i(x,x/\varepsilon)\frac{\partial v_{lin,j}^H(x)}{\partial x_i},\ x\in W(K_{\delta_j}),$$

where $\varepsilon\chi^i=\eta^i$ and $\chi^i(x,x/\varepsilon)=\chi^i(x,y)$ are defined in (23) (a similar representation holds for $v_j^h(x)$ with $\varepsilon\chi^{i,h}=\eta^{i,h}$). Assuming χ^i is smooth we obtain by the chain rule

$$\|D^\alpha(\varepsilon\chi^i)\|_{L^\infty(K_{\delta_j})}\le C\ \varepsilon^{-|\alpha|+1},\ \alpha\in\mathbb{N}^d, \tag{53}$$

where C is independent of ε.

Lemma 10 *Let $v^H,w^H\in V^p(\Omega,\mathcal{T}_H)$. Let $u^H,\bar{u}^H$ be the solutions of (37) and (47), respectively, with the same coupling condition (either (33) or (34)) for the micro problem (35). Assume that for $i=1,\ldots,d$, and for all $x_{j,K}\in\Omega$ such that $K_\delta(x_{j,K})\subset\Omega$, $\eta^i(x)\in H^{q+1}(K_\delta(x_{j,K}))$. Assume further that for $|\alpha|=q+1$, $\|D^\alpha\eta^i\|_{L^\infty(K_\delta(x_{j,K}))}\le C\ \varepsilon^{-|\alpha|+1}$ with a constant C independent of $x_{j,K}\in\Omega$ and $\delta>0$. Then*

$$\left|\bar{B}_H(v^H,w^H)-B_H(v^H,w^H)\right|\le C\ \left(\frac{h}{\varepsilon}\right)^{2q}\|\nabla v^H\|_{L^2(\Omega)}\|\nabla w^H\|_{L^2(\Omega)}. \tag{54}$$

Proof. We have (using the notation K_{δ_j} for $K_\delta(x_{j,K})$)

$$\begin{aligned}
&|B_H(v^H,w^H)-\bar{B}_H(v^H,w^H)| \qquad (55)\\
&=\left|\sum_{K\in\mathcal{T}_H}\sum_{j=1}^J\frac{\omega_{j,K}}{|K_{\delta_j}|}\left(\int_{K_{\delta_j}}a^\varepsilon(x)\nabla v_j\cdot\nabla w_j dx-\int_{K_{\delta_j}}a^\varepsilon(x)\nabla v_j^h\cdot\nabla w_j^h dx\right)\right|\\
&=\Big|\sum_{K\in\mathcal{T}_H}\sum_{j=1}^J\frac{\omega_{j,K}}{|K_{\delta_j}|}\int_{K_{\delta_j}}a^\varepsilon(x)\nabla(v_j-v_j^h)\cdot\nabla w_j dx\\
&-\int_{K_{\delta_j}}a^\varepsilon(x)\nabla v_j^h\cdot\nabla(w_j^h-w_j)dx\Big|.
\end{aligned}$$

We observe that the first member of the last line of (55) is zero since $(v_j - v_j^h) \in W(K_{\delta_j})$ (here we used the symmetry of a^ε). Using the same argument, replacing v_j^h by $v_j^h - v_j$ in the second expression and using the boundedness of a^ε we obtain

$$\begin{aligned}&|B_H(v^H, w^H) - \bar{B}_H(v^H, w^H)| \\ &\leq C \sum_{K\in\mathcal{T}_H} \sum_{j=1}^{J} \frac{\omega_{j,K}}{|K_{\delta_j}|} \|\nabla v_j^h - \nabla v_j\|_{L^2(K_{\delta_j})} \|\nabla w_j^h - \nabla w_j\|_{L^2(K_{\delta_j})}. \end{aligned} \tag{56}$$

Next, using the expression (52), the regularity assumption on η^i and standard FE approximation estimates [25, Thm. 3.2.2] we obtain

$$\begin{aligned}\|\nabla v_j^h - \nabla v_j\|_{L^2(K_{\delta_j})} &\leq C \max_i \|\nabla \eta^{i,h} - \nabla \eta^i\|_{L^2(K_{\delta_j})} |\nabla v_{lin,j}^H| \\ &\leq Ch^q |\eta^i|_{H^{q+1}(K_{\delta_j})} |\nabla v_{lin,j}^H| \leq C \left(\frac{h}{\varepsilon}\right)^q \sqrt{|K_{\delta_j}|} |\nabla v_{lin,j}^H| \\ &\leq C \left(\frac{h}{\varepsilon}\right)^q \|\nabla v_{lin}^H\|_{L^2(K_{\delta_j})},\end{aligned}$$

where C is independent of K_{δ_j} and $|\cdot|_{H^q(K_{\delta_j})}$ denotes the usual semi-norm in the Sobolev space $H^q(K_{\delta_j})$. Using a similar estimate for the second term of (56) we obtain the claimed estimate (54) by observing that

$$\begin{aligned}&|B_H(v^H, w^H) - \bar{B}_H(v^H, w^H)| \\ &\leq C \left(\frac{h}{\varepsilon}\right)^{2q} \sum_{K\in\mathcal{T}_H} \sum_{j=1}^{J} \omega_{j,K} |\nabla v^H(x_{j,K})| |\nabla w^H(x_{j,K})| \qquad (57) \\ &\leq C \left(\frac{h}{\varepsilon}\right)^{2q} \left(\sum_{K\in\mathcal{T}_H} \sum_{j=1}^{J} \omega_{j,K} |\nabla v^H(x_{j,K})|^2\right)^{1/2} \left(\sum_{K\in\mathcal{T}_H} \sum_{j=1}^{J} \omega_{j,K} |\nabla w^H(x_{j,K})|^2\right)^{1/2} \qquad (58) \\ &\leq C \left(\frac{h}{\varepsilon}\right)^{2q} \|\nabla v^H\|_{L^2(\Omega)} \|\nabla w^H\|_{L^2(\Omega)}, \qquad (59)\end{aligned}$$

where we used the Cauchy-Schwarz inequality, the identity $\nabla v_{lin}^H(x) \equiv \nabla v^H(x_{j,K})$ for $x \in K$ and the assumptions on the QF. □

Remark 11 *Without the symmetry assumption on a^ε we obtain the weaker estimate*

$$\left|\bar{B}_H(v^H, w^H) - B_H(v^H, w^H)\right| \leq C \left(\frac{h}{\varepsilon}\right)^q \|\nabla v^H\|_{L^2(\Omega)} \|\nabla w^H\|_{L^2(\Omega)}.$$

Remark 12 *If we denote by $M = \dim S^1(K_\delta, \mathcal{T}_h)$ (degrees of freedom (DOF)) and suppose that $\delta = C\varepsilon$ (with C a moderate constant independent of ε) then the mesh size of the micro FE space on K_δ (of measure $|K_\delta| = \delta^d$) is given by $h = C\varepsilon M^{-\frac{1}{d}}$. Therefore, the quantity h/ε in (48) or (54) is independent of ε and we can express it as $CM^{-\frac{1}{d}}$, which emphasizes that it depends only on the DOF of $S^1(K_\delta, \mathcal{T}_h)$. The same is true for $S^q(K_\delta, \mathcal{T}_h)$ with obvious changes to factor out the additional local DOF.*

Modeling error. The last contribution to the error of the FE-HMM approximation of the multiscale elliptic problem is the so-called modeling error, i.e., the difference $\|u^{0,H} - \bar{u}^H\|$, where $u^{0,H}$ is the solution of the problem (44) and $\bar{u}^H$ is the solution of the problem (47). Here some knowledge about the specific form of the small scales is needed in order to obtain error bounds. We suppose in what follows that $a^\varepsilon(x) = a(x, x/\varepsilon) = a(x, y)$ Y-periodic in y, where $Y = (0,1)^d$. In this situation the homogenized tensor $a^0(x)$ is given by (22).

Remark 13 *If an explicit form $a(x, x/\varepsilon)$ of the tensor $a^\varepsilon(x)$ is available it can be advantageous to replace the bilinear form (46) by*

$$\widetilde{B}_H(v^H, w^H) = \sum_{K\in\mathcal{T}_H} \sum_{j=1}^{J} \frac{\omega_{j,K}}{|K_\delta(x_{j,K})|} \int_{K_\delta(x_{j,K})} a(x_{j,K}, x/\varepsilon)\nabla v_j \cdot \nabla w_j dx, \tag{60}$$

where we collocate the slow variable of $a^\varepsilon(x)$ at the nodes of the quadrature formula and where v_j and w_j are the exact solutions of the cell problem (35) with $a^\varepsilon(x)$ replaced by $a(x_{j,K}, x/\varepsilon)$. Define $\tilde{u}^H$ to be the solution of

$$\widetilde{B}_H(\tilde{u}^H, v^H) = F(v^H). \tag{61}$$

Similarly, we can modify the FE-HMM itself by collocating, as done above, the slow variables in (31) and in (35).

We will see below that for the FE-HMM the periodic coupling (33) is optimal for a (non uniformly) periodic tensor, and the minimal computational cost is achieved by setting $\delta = \varepsilon$, i.e., $K_\delta(x_{j,K}) = K_\varepsilon(x_{j,K})$. The following Proposition is based on a result first obtained in [2, Equ. (50)] (see also [5, Appendix A.1]).

Proposition 14 *Let $a^\varepsilon(x) = a(x, x/\varepsilon) = a(x, y)$ Y-periodic in y, and $\bar{u}^H, \tilde{u}^H$ be the solutions of (47) and (61), respectively, where exact micro functions are used in both bilinear forms with a periodic coupling condition (33). Suppose further that $\delta/\varepsilon \in \mathbb{N}$ and that $a^\varepsilon_{ij}(x, y) \in W^{1,\infty}(\bar{\Omega}, L^\infty(Y))$ $\forall i, j = 1, \ldots d$. Then,*

$$u^{0,H} = \tilde{u}^H \quad \textit{and} \quad \|u^{0,H} - \bar{u}^H\|_{H^1(\Omega)} \leq C\varepsilon, \tag{62}$$

where $u^{0,H}$ is the solution of (44).

Proof. Let us first assume $\delta = \varepsilon$. Observe that v_j, the exact solution of the cell problem (35) with $a^\varepsilon(x)$ replaced by $a(x_{j,K}, x/\varepsilon)$, is given by

$$v_j = v^H_{lin,j}(x) + \sum_{i=1}^{d} \varepsilon\chi^i(x_{j,K}, x/\varepsilon)\frac{\partial v^H_{lin,j}(x)}{\partial x_i},$$

and similarly for w_j. We then compute

$$\begin{aligned}
&\frac{1}{|K_\varepsilon(x_{j,K})|}\int_{K_\varepsilon(x_{j,K})} a(x_{j,K},x/\varepsilon)\nabla\left(v^H_{lin,j}(x)+\sum_{i=1}^d \varepsilon\chi^i(x_{j,K},x/\varepsilon)\frac{\partial v^H_{lin,j}(x)}{\partial x_i}\right)\\
&\quad\cdot\nabla\left(w^H_{lin,j}(x)+\sum_{i=1}^d \varepsilon\chi^i(x_{j,K},x/\varepsilon)\frac{\partial w^H_{lin,j}(x)}{\partial x_i}\right)dx\\
&=\frac{1}{|K_\varepsilon(x_{j,K})|}\int_{K_\varepsilon(x_{j,K})} a(x_{j,K},x/\varepsilon)\left(I+\nabla_y\boldsymbol{\chi}(x_{j,K},x/\varepsilon)\right)\nabla v^H_{lin,j}(x)\cdot\nabla w^H_{lin,j}(x)dx\\
&=a^0(x_{j,K})\nabla v^H(x_{j,K})\cdot\nabla w^H(x_{j,K}) \qquad (63)
\end{aligned}$$

where we used the notation $\nabla\boldsymbol{\chi}=(\nabla\chi^1,\dots,\nabla\chi^d)$, that $\chi^i(x_{j,K},y)$ $(y=x/\varepsilon)$ is a solution of (23) and the identity $\nabla v^H_{lin,j}(x)\equiv\nabla v^H(x_{j,K})$ for $x\in K$. Thus $\widetilde{B}_H(\cdot,\cdot)=B_{0,H}(\cdot,\cdot)$ (see (43)) and the first claim of the lemma is proved. For the second inequality, we have to estimate

$$\begin{aligned}
&|\bar{B}_H(v^H,w^H)-\widetilde{B}_H(v^H,w^H)|\\
&=\left|\sum_{K\in\mathcal{T}_H}\sum_{j=1}^J\frac{\omega_{j,K}}{|K_\varepsilon(x_{j,K})|}\int_{K_\varepsilon(x_{j,K})}\left(a(x_{j,K},x/\varepsilon)-a(x,x/\varepsilon)\right)\nabla v_j\cdot\nabla w_j dx\right|\\
&\le C\varepsilon\sum_{K\in\mathcal{T}_H}\sum_{j=1}^J\omega_{j,K}|\nabla v^H(x_{j,K})||\nabla w^H(x_{j,K})|\\
&\le C\varepsilon\left(\sum_{K\in\mathcal{T}_H}\sum_{j=1}^J\omega_{j,K}|\nabla v^H(x_{j,K})|^2\right)^{1/2}\left(\sum_{K\in\mathcal{T}_H}\sum_{j=1}^J\omega_{j,K}|\nabla w^H(x_{j,K})|^2\right)^{1/2}\\
&\le C\varepsilon\|\nabla v^H\|_{L^2(\Omega)}\|\nabla w^H\|_{L^2(\Omega)}, \qquad (64)
\end{aligned}$$

where we used (40), the Cauchy-Schwarz inequality and the assumption on the tensor a^ε and on the QF. Using an inequality similar to (50) gives the second assertion of the lemma and the proof is complete by noting that the arguments remain unchanged when $\delta>\varepsilon$ and $\delta/\varepsilon\in\mathbb{N}$. □

Remark 15 *For a tensor of the form $a^\varepsilon(x)=a(x,x/\varepsilon)$ with explicit separation between fast and slow scales we obtain, by collocating the slow variables in the FE-HMM (see Remark 13) and in view of Proposition 8 (replacing $\bar{u}^H$ by $\tilde{u}^H$), the estimate $\|u^{0,H}-u^H\|_{H^1(\Omega)}\le C\left(\frac{h}{\varepsilon}\right)^{2q}$.*

The Proposition 14 and Remark 15 show that for a periodic tensor the periodic boundary conditions are optimal. No modeling error occurs for the FE-HMM if a bilinear form with collocated slow variables is used. In practice, even for periodic problems, it may happen that the size of the sampling domain (the period) is only approximatively known. Then it is of interest to study the case $\delta>\varepsilon$, with δ/ε non integer. In this situation, boundary layers occur and we have for $\delta>\varepsilon$ [33, Thm. 1.2]

$$\|u^{0,H}-\bar{u}^H\|_{H^1(\Omega)}\le C(\delta+\frac{\varepsilon}{\delta}), \qquad (65)$$

where $\bar{u}^H$ is the solution of (47) with Dirichlet boundary conditions (34) and $u^{0,H}$ is the solution of (31).

A-priori error estimate: convergence Theorems. As explained in Section 3.3 (see (38), the fully discrete error is made of three contributions: e_{MAC} (micro error Proposition 7) e_{MIC} (macro error Proposition 8), e_{MOD} (micro error Proposition 14 and (65)). Collecting these results, we obtain the fully discrete analysis for the FE-HMM.

Theorem 16 (Fully discrete analysis: $\delta/\varepsilon \in \mathbb{N}$, periodic coupling (33))
Suppose that the assumptions of Theorem 6 and Propositions 7,8 and 14 hold. Then

$$\|u^0 - u^H\|_{H^1(\Omega)} \leq C\left(H^p + \left(\frac{h}{\varepsilon}\right)^{2q} + \varepsilon\right), \tag{66}$$

$$\|u^0 - u^H\|_{L^2(\Omega)} \leq C\left(H^{p+1} + \left(\frac{h}{\varepsilon}\right)^{2q} + \varepsilon\right), \tag{67}$$

where u^0 is the solution of (21) and u^H the solution of (37).
Under the same assumptions but with a collocated slow variable in the tensor a^ε (see Remark 15) we obtain

$$\|u^0 - u^H\|_{H^1(\Omega)} \leq C\left(H^p + \left(\frac{h}{\varepsilon}\right)^{2q}\right), \tag{68}$$

$$\|u^0 - u^H\|_{L^2(\Omega)} \leq C\left(H^{p+1} + \left(\frac{h}{\varepsilon}\right)^{2q}\right). \tag{69}$$

Corollary 1 *Suppose that the assumption of Theorem 16 hold. Then*

$$\|u^\varepsilon - u^H\|_{L^2(\Omega)} \leq C\left(H^{p+1} + \left(\frac{h}{\varepsilon}\right)^{2q} + \varepsilon\right), \tag{70}$$

where u^ε is the solution of (16) and u^H the solution of (37).

Proof. The result follows from the above Theorem and the estimate (24). □

Theorem 17 (Fully discrete analysis: $\delta > \varepsilon, \delta/\varepsilon \notin \mathbb{N}$, Dirichlet coupling (34))
Suppose that the assumptions of Theorem 6, Propositions 7,8 and (65) hold. Then

$$\|u^0 - u^H\|_{H^1(\Omega)} \leq C\left(H^p + \left(\frac{h}{\varepsilon}\right)^{2q} + \delta + \frac{\varepsilon}{\delta}\right), \tag{71}$$

$$\|u^0 - u^H\|_{L^2(\Omega)} \leq C\left(H^{p+1} + \left(\frac{h}{\varepsilon}\right)^{2q} + \delta + \frac{\varepsilon}{\delta}\right), \tag{72}$$

where u^0 is the solution of (21) and u^H the solution of (37).

Corollary 2 *Suppose that the assumption of Theorem 17 hold. Then*

$$\|u^\varepsilon - u^H\|_{L^2(\Omega)} \leq C\left(H^{p+1} + \left(\frac{h}{\varepsilon}\right)^{2q} + \delta + \frac{\varepsilon}{\delta}\right), \tag{73}$$

where u^ε is the solution of (16) and u^H the solution of (37).

Proof. The result follows from the above Theorem and the estimate (24). □

The above theorems fully describe the convergence of the FE-HMM to the effective (homogenized) solution of the multiscale problem (16). The sampling domains size and the coupling conditions are responsible for the modeling error. Once chosen, an appropriate micro and macro mesh refinement has to be implemented in order to have the best possible convergence rate for the minimal computational cost. More precisely, Theorems 16 and 17 show that micro and macro mesh have to be refined simultaneously and give precise speed at which this need to be done.

Recovery of the homogenized tensor. We explain here how an approximation of the homogenized tensor $a^0(x)$ can be computed during the elementwise assembly process of the FE-HMM. We assume that $a^\varepsilon(x) = a(x, x/\varepsilon) = a(x, y)$ Y-periodic in y and restrict ourself to piecewise linear simplicial macro and micro FE (higher order approximations can be obtained with higher order micro FE following the lines of the discussion below). In this situation $u^H = u^H_{lin}$ we choose periodic constraints (33) in the FE-HMM and sampling domains with $\delta = \varepsilon$. Consider a triangle $K \in \mathcal{T}_H$, and $V_K \subset V^1(\Omega, \mathcal{T}_H)$ the collection of nodal basis functions associated with the vertices of K. Remember that the following expression is computed during the FE-HMM assembly process

$$\frac{1}{|K_{\varepsilon(x_K)}|}\int_{K_{\varepsilon(x_K)}} a(x, x/\varepsilon)\nabla\varphi_i^h \cdot \nabla\varphi_j^h dx, \tag{74}$$

where φ_i^h or φ_j^h are solutions of (35) constrained by $\varphi_i^H, \varphi_j^H \in V_K$. Consider the affine mapping (C^1 diffeomorphism) $F_K : \hat{K} \to K$, $F_K(\hat{x}) = x$, which maps the reference simplex $\hat{K} = \{\hat{x} \in \mathbb{R}^d; \hat{x}_i > 0, \sum_{i=1}^d \hat{x}_i < 1\}$ onto K. The nodal basis of the reference simplex is defined by $\hat{\varphi}_i^H = \hat{x}_i,\ i = 1, \ldots, d,\ \hat{\varphi}_0^H = 1 - \sum_{i=1}^d \hat{x}$. We order the nodal basis of V_K so that $\varphi_i^H(F_K(\hat{x})) = \hat{\varphi}_i^H(\hat{x}),\ i = 0, \ldots, d$ and define the matrix $M_K^h \in \mathbb{R}^{d\times d}$ by $(M_K^h)_{ij} = (B(\varphi_i^H, \varphi_j^H))_{ij},\ i, j = 1, \ldots, d$. We also consider two matrices obtained similarly as above. The first, denoted by $(\widetilde{M}_K^h)_{ij}$ is obtained with a collocated bilinear form for the FE-HMM, the second denoted by $(\widetilde{M}_K)_{ij}$ is obtained with a collocated bilinear form and exact micro solutions (see Remark 13).

Theorem 18 *Define $(a^{0,h}(x_K))_{ij} = (M_K^h)_{ij}$. Then $(a^{0,h}(x_K))_{ij}$ is an approximation of the homogenized tensor $a^0(x)$ (see (22)) at the integration point x_K and we have*

$$\left|a^{0,h}(x_K)_{ij} - a^0(x_K)_{ij}\right| \leq C\left(\left(\frac{h}{\varepsilon}\right)^2 + \varepsilon\right), \tag{75}$$

where h is the meshsize of the micro FEM used in (35). If a collocated bilinear form is used for the FE-HMM (see Remark 13) then the estimate (75) can be improved as follows

$$\left|a^{0,h}(x_K)_{ij} - a^0(x_K)_{ij}\right| \leq C\left(\frac{h}{\varepsilon}\right)^2. \tag{76}$$

Proof. In view of equality (63) we have

$$(\widetilde{M}_K)_{ij} = a^0(x_K)\nabla\varphi_i^H(x_K)\cdot\nabla\varphi_j^H(x_K). \tag{77}$$

A simple change of variables (recall that φ_i^H is a nodal (piecewise linear) basis function) gives

$$J_K^T(\widetilde{M}_K)_{ij}J_K = a^0(x_K)e_i\cdot e_j, \tag{78}$$

where $(e_j)_{j=1}^d$ is the canonical basis of $\mathbb{R}^d$ and J_K is the Jacobian matrix of F_K. We first assume that we use a collocated bilinear form in the FE-HMM. Then, similarly as above, we see that $(\widetilde{M}_K^h)_{ij} = a^{0,h}(x_K)\nabla\varphi_i^H(x_K)\cdot\nabla\varphi_j^H(x_K)$ and $J_K^T(\widetilde{M}_K^h)_{ij}J_K = a^{0,h}(x_K)e_i\cdot e_j$. We then obtain

$$\left|a^{0,h}(x_K)_{ij} - a^0(x_K)_{ij}\right| \leq C\left(\frac{h}{\varepsilon}\right)^2,$$

by noting that

$$\left|a^{0,h}(x_K)_{ij} - a^0(x_K)_{ij}\right| = \left|\frac{1}{|K_{\varepsilon(x_K)}|}\int_{K_{\varepsilon(x_K)}} a(x_K, x/\varepsilon)\left(\nabla\tilde{\varphi}_i^h\cdot\nabla\tilde{\varphi}_j^h - \nabla\tilde{\varphi}_i\cdot\nabla\tilde{\varphi}_j dx\right)\right|,$$

and by using Lemma 10 to bound the right hand side of the above expression. This proves the second claim of the theorem. The first claim of the theorem follows from a triangle inequality by using an estimation similar to (64) (see Proposition 14) for a macro element K. □

Remark 19 *For a nonsymmetric tensor a^ε, one should define $(a^{0,h}(x_K))_{ij} = (M_K^h)_{ji}$ and following the above proof (with obvious changes) we only obtain a linear convergence rate in $\left(\frac{h}{\varepsilon}\right)$ (see also Remark 12).*

Numerical examples. We present here several numerical examples to give some insight on the sharpness of the bounds obtained above. We thus deliberately choose very simple multiscale problems to be able to compute reference solutions with high precision and to know the optimal size of the sampling domains for the FE-HMM.

We first consider the following multiscale problem [6]

$$\begin{aligned} -\nabla\cdot(a^\varepsilon(x)\nabla u^\varepsilon) &= f(x) \quad \text{in } \Omega = (0,1)^2, \\ u^\varepsilon|_{\Gamma_D} &= 0 \quad \text{on } \Gamma_D := \{x_1 = 0\}\cup\{x_1 = 1\}, \\ n\cdot(a^\varepsilon(x)\nabla u^\varepsilon)\,|_{\Gamma_N} &= 0 \quad \text{on } \Gamma_N := \partial\Omega\backslash\Gamma_D, \end{aligned} \tag{79}$$

where $a^\varepsilon = a(x/\varepsilon) = a(y) = (\cos 2\pi y_1 + 2)I$, $y = (y_1, y_2) \in Y = (0,1)^2$, and $f(x) \equiv 1$. The

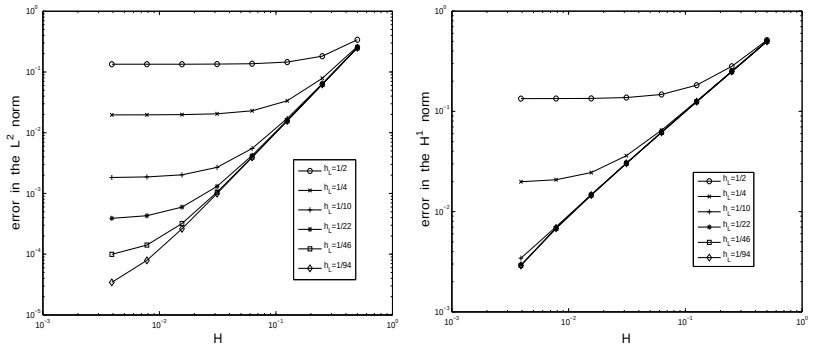


Figure 13: Error against the homogenized solution u^0 for the FE-HMM applied to problem (79) with macromesh refinement for fixed micro mesh $h_L = 1/2, 1/4, 1/10, 1/22, 1/46$.

exact solution as well as the homogenized tensor can be derived analytically

$$u^\varepsilon = -\int_0^{x_1} \frac{t}{a(t/\varepsilon)}dt + \frac{\int_0^1 \frac{t}{a(t/\varepsilon)}dt}{\int_0^1 \frac{1}{a(t/\varepsilon)dt}} \int_0^{x_1} \frac{1}{a(t/\varepsilon)}dt, \quad a^0 = \begin{pmatrix} (\int_0^1 \frac{1}{a(y_1)}dy_1)^{-1} & 0 \\ 0 & 2 \end{pmatrix}.$$

We can therefore compute a reference solution for the fine scale solution and for the homogenized solution with high precision. The reference solution for u^ε is computed with the above integral formula (with a very precise numerical integration scheme). The homogenized solution is a quadratic polynomial obtained from the solution of (79) with a^0 instead of a^ε and can be easily computed. In Figure 13 we report numerical results for the problem (79) solved with the FE-HMM. We choose piecewise linear macro and micro FE spaces and periodic coupling. If we further choose $\delta = \varepsilon$ for the sampling domain and a "collocated bilinear form" (see Remark 15), Theorem 16 gives us the following a priori convergence rates

$$\|u^0 - u^H\|_{H^1(\Omega)} \leq C\left(H + \left(\frac{h}{\varepsilon}\right)^2\right), \ \|u^0 - u^H\|_{L^2(\Omega)} \leq C\left(H^2 + \left(\frac{h}{\varepsilon}\right)^2\right).$$

We set $h = \varepsilon/L$ for the micromesh, $h_L = h/\varepsilon = 1/L$ and $H_M = 1/M$ for the macromesh. Denoting by N_{mac} the macro DOF and by N_{mic} the micro DOF, the above rates of convergence show that

$$N_{mic} = N_{mac} \quad (L^2 \text{ norm}), \qquad N_{mic} = \sqrt{N_{mac}} \quad (H^1 \text{ norm}),$$

i.e., $h_L = H_M$ in the L^2 norm and $h_L = \sqrt{H_M}$ for the H^1 norm are the best refinement strategies for optimal convergence rates with minimal computational cost. We thus obtain a complexity of $\mathcal{O}(N_{mac} \cdot N_{mic}) = \mathcal{O}(N_{mac}^{3/2})$ floating point operations for a linear (macro) convergence rate in the H^1 norm and $\mathcal{O}(N_{mac} \cdot N_{mic}) = \mathcal{O}(N_{mac}^2)$ floating point operations for a quadratic convergence rate in the L^2 norm. Here we assume that the cost (floating point operations) of the method is proportional to the total DOF (which is the case for example when using multigrid linear solver). We see in Figure 13 that the numerical results are in perfect agreement with the theoretical convergence rates. We compute the solution of problem (79) with successive macro grid refinement $H_M = 1/2, 1/4, 1/10, 1/22, 1/46$. The micro mesh h_L is kept fixed for each solid line in Figure 13 and is successively refined from

one solid line to the other. Optimal refinements clearly follow the ratio $h_L = H_M$ (L^2 norm) and $h_L = \sqrt{H_M}$ (H^1 norm). This demonstrates the sharpness of the a priori bounds. Similar results for piecewise bilinear FE (quadrilateral elements) are reported in [9].

We study next the modeling error and the influence of the choice of the boundary conditions by applying the FE-HMM to the same test problem but choosing deliberately sampling domains K_δ with $\delta/\varepsilon \notin \mathbb{N}$. We choose $K_\delta = 1.1\varepsilon$ and $K_\delta = (5/3)\varepsilon$ and compute the FE-HMM solution of problem (79) on macro meshes $H_M = 1/2, 1/4, 1/8, 1/16$, first with Dirichlet boundary conditions (34) then with periodic boundary conditions (33) for the micro solver. As we want to observe the influence of the coupling conditions and the size of the sampling domains, we solve the micro problem with a fine mesh in order to ensure that micro error are negligible. We see in Figure 14 that the choice of Dirichlet coupling conditions has an important impact on the quality of the approximation. This can be better seen in the L^2 norm as the macro error decreases more rapidly. Increasing the size of the sampling domain from $K_\delta = 1.1\varepsilon$ to $K_\delta = (5/3)\varepsilon$ improves the results. In Figure 15 we perform similar experi-

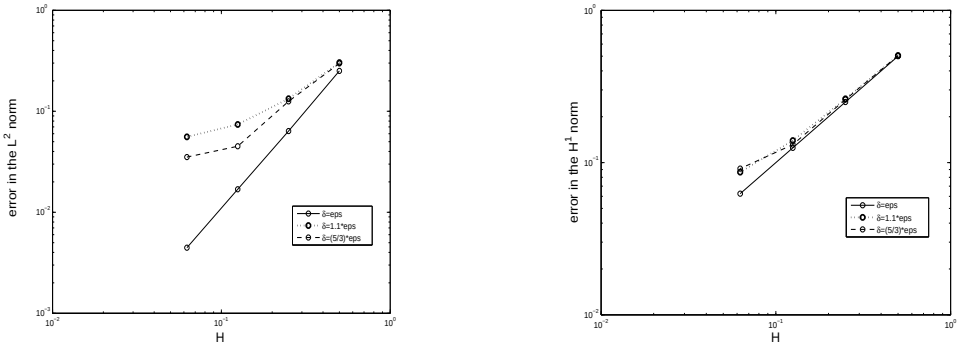


Figure 14: Error against the homogenized solution u^0 for the FEHMM applied to problem (79) with macromeshes $H_M = 1/2, 1/4, 1/8, 1/16$, Dirichlet coupling conditions and successively $\delta = 1.1\varepsilon$, $\delta = (5/3)\varepsilon$ (L^2 error (left picture), H^1 error (right picture)). The graph for $\delta = \varepsilon$ is obtained by the FE-HMM with (optimal) periodic boundary conditions. The micro mesh for all experiments is small enough to ensure negligible micro errors.

ments but this time with periodic boundary conditions. The results are much better and the influence of the non matching size of the sampling domains are much smaller than previously, with Dirichlet coupling condtions. The better performance of periodic boundary conditions for such type of multiscale problems (even with non-matching size of sampling domains) has been observed frequently, but a complete theoretical understanding and analysis is still to be done (see the related discussion and references in [65]).

Let us now apply these optimal refinement procedures to another problem

$$\begin{aligned} -\nabla \cdot (a^\varepsilon(x)\nabla u^\varepsilon) &= f(x) \quad \text{in } \Omega = (0,1)^2, \\ u^\varepsilon &= 0 \qquad \text{on } \partial\Omega, \end{aligned} \tag{80}$$

where

$$a^\varepsilon = \begin{pmatrix} 2 + \sin(2\pi(x_1/\varepsilon)) & 0 \\ 0 & 2 + \sin(2\pi(x_2/\varepsilon)) \end{pmatrix}.$$

We compute a reference homogenized solution using (22) and study the convergence in the L^2 and H^1 norm for decreasing macro meshes $H_M = 1/2, 1/4, 1/8, 1/16$ and we set the micro

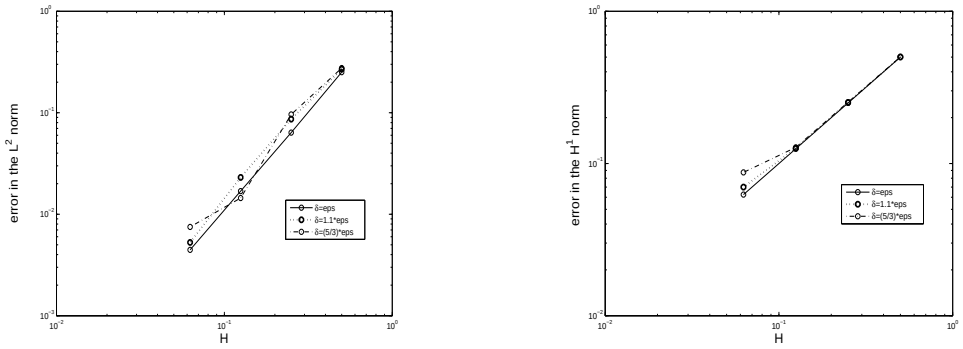


Figure 15: Error against the homogenized solution u^0 for the FE-HMM applied to problem (79) with macromeshes $H_M = 1/2, 1/4, 1/8, 1/16$, periodic coupling conditions and successively $\delta = 1.1\varepsilon$, $\delta = (5/3)\varepsilon$ (L^2 error (left picture), H^1 error (right picture)). The graph for $\delta = \varepsilon$ is obtained by the FE-HMM with (optimal) periodic boundary conditions. The micro mesh for all experiments is small enough to ensure negligible micro errors.

mesh to $h_L = H_M$ (L^2 norm) and $h_L = \sqrt{H_M}$ (H^1 norm). Piecewise linear FE are again used at macro and micro level for the FE-HMM. We see in Figure 16 the corresponding macro solution. In Figure 17 we observe that the expected (optimal) macro convergence rates are obtained when we follow the aforementioned micro-macro refinement strategy.

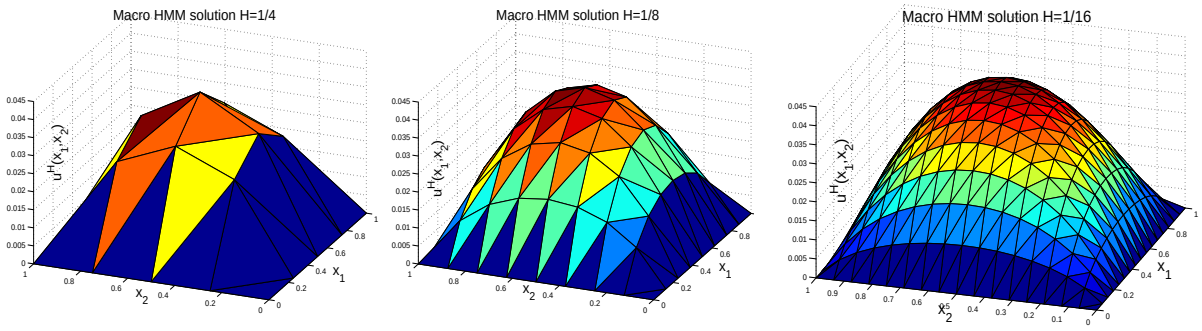


Figure 16: FE-HMM solution for problem (80) for decreasing macro mesh $H_M = 1/4, 1/8, 1/16$ (optimal corresponding micro mesh refinement).

3.3.3 Post-processing procedure: modeling and analysis

The results of Theorems 16 and 17 show that $u^H \to u^0$ in the H^1 norm and $u^H \to u^\varepsilon$ in the L^2 norm. As explained in Section 3.2, convergence $u^H \to u^\varepsilon$ or $u^0 \to u^\varepsilon$ does not occur in the H^1 norm since u^H or u^0 do not carry any information on the gradient of the oscillation occurring in u^ε. We need thus to correct the solutions u^H, u^0 by adding information on the small scale. Such a procedure has been described for the homogenized solution u^0 (see (26)). Computing numerically the corrector (25) for all $x \in \Omega$ is as costly as solving the original problem. For the FE-HMM one can use an idea first proposed in [59], although not in an HMM context. The known small scale solution in the sampling domain (35) computed during the assembly of the FE-HMM can be extended locally on the macro element K and added to u^H. The error in the slow variable of this corrector will be proportional to the size

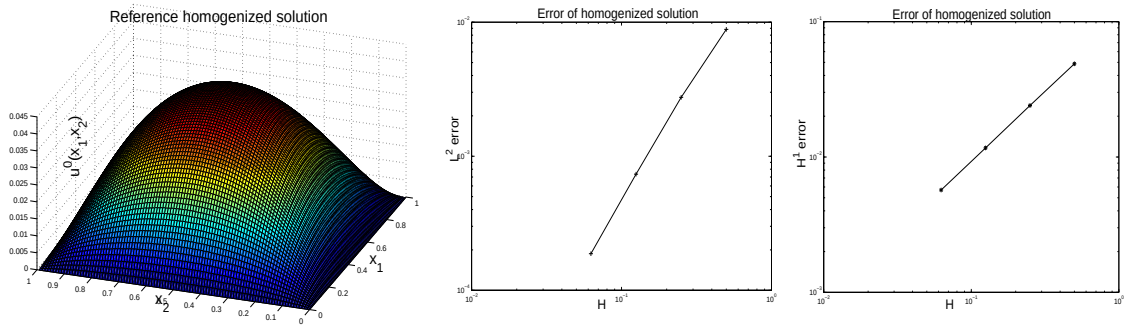


Figure 17: Reference solution for the homogenized solution of problem (80) (left picture) macro convergence rate with adapted (optimal) micro mesh refinement (middle picture L^2 norm, right picture H^1 norm).

of the macro triangle, while self similarity in the small scale justifies such an extension of the micro information.

We consider the function $(u^h - u^H)$, available in $K_\delta(x_{j,K}) \subset K$ and extend it periodically on the whole element K. We set

$$u_{p,\varepsilon}(x)|_K = u^H(x) + \left(u^h - u^H\right)\left(x - [x]_{K_\delta(x_{j,K})}\right) \text{ for } x \in K \in \mathcal{T}_H, \tag{81}$$

where for $x \in \mathbb{R}^d$, $[x]_{K_\delta(x_{j,K})}$ denotes the unique combination $\delta \sum_{i=1}^d b_i e_i$, where $b_i \in \mathbb{Z}$ and $(e_i)_{i=1}^d$ is the canonical basis of $\mathbb{R}^d$, such that $(x - [x]_{K_\delta(x_{j,K})}) \in K_\delta(x_{j,K})$ (see Figure 18). Since $u_{p,\varepsilon}$ can be discontinuous across the macro elements K, we define a broken H^1 norm by

$$\|u\|_{\bar{H}^1(\Omega)} := \Big(\sum_{K \in \mathcal{T}_H} \|\nabla u\|^2_{L^2(K)}\Big)^{1/2}. \tag{82}$$

In what follows we assume that $a^\varepsilon(x) = a(x, x/\varepsilon) = a(x, y)$ Y-periodic in y and restrict

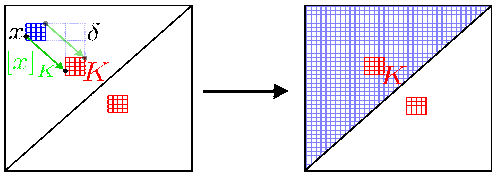


Figure 18: Post-processing procedure: the known small scale solution on the sampling domain (red domain) is extended on the macro element K (blue domain).

ourself to piecewise linear simplicial macro and micro FE. In this situation $u^H = u^H_{lin}$ we choose periodic constraints (33) in the FE-HMM and sampling domains with $\delta = \varepsilon$. Then $u_{p,\varepsilon}(x)$ restricted to each macro element K can be written as

$$u_{p,\varepsilon}(x)|_K = u^H(x) + \sum_{i=1}^d \varepsilon \chi^{i,h}(x - [x]_{K_\varepsilon(x_K)}, x/\varepsilon) \frac{\partial u^H(x)}{\partial x_i}, \ x \in K, \tag{83}$$

where $u_{p,\varepsilon}(x) - u^H(x) = u^h(x)$ and $u^h(x)$ is solution of (35). Let $\bar{u}_{p,\varepsilon}(x) = \bar{u}^H(x) + \left(u - \bar{u}^H\right)\left(x - [x]_{K_\delta(x_{j,K})}\right)$ where $\bar{u}^H(x)$ is the solution of the semi-discrete problem (47) and

$u(x)$ the corresponding micro function (see (46)). Then $\bar{u}_{p,\varepsilon}(x)$ can be written as

$$\bar{u}_{p,\varepsilon}(x)|_K = \bar{u}^H(x) + \sum_{i=1}^{d} \varepsilon \chi^i(x - [x]_{K_\varepsilon(x_K)}, x/\varepsilon) \frac{\partial \bar{u}^H(x)}{\partial x_i}, \ x \in K. \tag{84}$$

Following the line of Proposition 8 (for $q = 1$) we obtain

$$\|u_{p,\varepsilon}(x) - \bar{u}_{p,\varepsilon}(x)\|_{\bar{H}^1(\Omega)} \leq C \left(\frac{h}{\varepsilon}\right). \tag{85}$$

For the analysis, we need to consider a post-processing procedure defined upon the (macro and micro) solutions of the collocated bilinear form. Let $\tilde{u}_{p,\varepsilon}(x) = \tilde{u}^H(x) + \left(\tilde{u} - \tilde{u}^H\right)(x - [x]_{K_\varepsilon(x_{j,K})})$ where $\tilde{u}^H(x)$ is the solution of the semi-discrete problem (61) and $\tilde{u}(x)$ the corresponding micro function (see (60)). Then $\tilde{u}_{p,\varepsilon}(x)$ can be written as

$$\tilde{u}_{p,\varepsilon}(x)|_K = \tilde{u}^H(x) + \sum_{i=1}^{d} \varepsilon \tilde{\chi}^i(x_K, x/\varepsilon) \frac{\partial \tilde{u}^H(x)}{\partial x_i}, \ x \in K. \tag{86}$$

Lemma 20 *Let $\bar{u}_{p,\varepsilon}(x)$ be given by (84) and $\tilde{u}_{p,\varepsilon}(x)$ be given by (86). Suppose that the assumption of Proposition 14 hold. Then*

$$\|\bar{u}_{p,\varepsilon}(x) - \tilde{u}_{p,\varepsilon}\|_{\bar{H}^1(\Omega)} \leq C\varepsilon. \tag{87}$$

Proof. The proof follows from Proposition 14. □

Theorem 21 *Let $u^\varepsilon(x)$ be the solution of (16) and $u_{p,\varepsilon}(x)$ given by (83). Suppose that the assumptions of Theorem 6 and Proposition 7 (for $p = 1$), Proposition 8 (for $q = 1$) and Proposition 14 hold. Suppose further that for $x \in \bar{\Omega} \to D^\alpha \chi^j(x, \cdot)$ is Lipschitz continuous for $|\alpha| = 1$. Then*

$$\|u^\varepsilon - u_{p,\varepsilon}\|_{\bar{H}^1(\Omega)} \leq C(H + \frac{h}{\varepsilon} + \sqrt{\varepsilon}). \tag{88}$$

Proof. We decompose the error as follows

$$\begin{aligned} \|u^\varepsilon - u_{p,\varepsilon}(x)\|_{\bar{H}^1(\Omega)} &\leq \|u^\varepsilon - (u^0 + \varepsilon u_1(x, x/\varepsilon)\|_{\bar{H}^1(\Omega)} + \|(u^0 + \varepsilon u_1(x, x/\varepsilon) - \tilde{u}_{p,\varepsilon}(x)\|_{\bar{H}^1(\Omega)} \\ &+ \|\tilde{u}_{p,\varepsilon}(x) - u_{p,\varepsilon}(x)\|_{\bar{H}^1(\Omega)} \\ &= I_1 + I_2 + I_3. \end{aligned}$$

In view of (26) we have $I_1 \leq C\sqrt{\varepsilon}$. Using (87) and (85) we obtain $I_3 \leq C\left(\left(\frac{h}{\varepsilon}\right) + \varepsilon\right)$. Finally we have $I_2 \leq C(H + \varepsilon)$. Indeed,

$$\sum_{K \in \mathcal{T}_H} \left\|\nabla\left(\left(u^0 + \varepsilon u_1(x, x/\varepsilon)\right) - \tilde{u}_{p,\varepsilon}(x)\right)\right\|^2_{L^2(K)} \leq \sum_{K \in \mathcal{T}_H} \left\|\nabla(u^0 - \tilde{u}^H)\right\|^2_{L^2(K)} +$$

$$\sum_{K \in \mathcal{T}_H} \left\| \sum_{j=1}^{d} \nabla(\varepsilon \chi^j(x, x/\varepsilon)) \left(\frac{\partial u^0}{\partial x_j} - \frac{\partial \tilde{u}^H}{\partial x_j}\right) \right\|^2_{L^2(K)} +$$

$$\sum_{K \in \mathcal{T}_H} \left\| \varepsilon \sum_{j=1}^{d} \nabla\left(\chi^j(x_K, x/\varepsilon) - \chi^j(x, x/\varepsilon)\right) \frac{\partial \tilde{u}^H}{\partial x_j} \right\|^2_{L^2(K)}.$$

The first two terms are bounded by CH in view of (62) and (45) (for $p = 1$). Expanding the last term and using the assumption on $D^\alpha \chi^j$, we can bound the last term by $C(H + \varepsilon)$. Collecting the estimates for I_1, I_2, I_3 gives (88). □

Numerical examples. We present here some numerical results for the described post-processing procedure allowing to obtain an energy approximation of the small scale solution of a multiscale problem. As done previously, we set $h = \varepsilon/L$ for the micromesh, $h_L = h/\varepsilon = 1/L$ and $H_M = 1/M$ for the macromesh. The estimate (87) shows that $h_L = H_M$ is the best refinement strategy for optimal convergence with minimal computational cost. Denoting by N_{mac} the macro DOF and setting the micro DOF as $N_{mic} = N_{mac}$, we obtain a complexity of $\mathcal{O}(N_{mac} \cdot N_{mic}) = \mathcal{O}(N_{mac}^2)$ floating point operations for a linear (macro) convergence rate in the broken H^1 norm. We consider the problem (80) with $\varepsilon = 10^{-2}$. Having obtained a macro solution u^H with the FE-HMM (see Figure 16), we extend the stored micro solution available in the sampling domain K_ε over the whole macro element K as explained in (81). We present in Figure 19 (compare with Figure 16) the corresponding reconstructed solution. In Figure 20 we observe that the expected convergence rate predicted by Theorem 21 is obtained until a certain threshold $\simeq 10^{-1}$ which corresponds to $\sqrt{\varepsilon}$ as predicted in the error bounds (87).

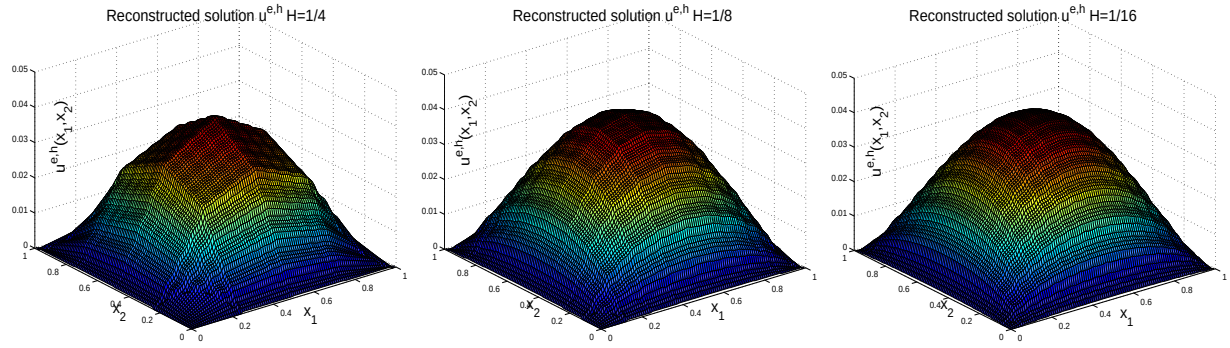


Figure 19: Reconstructed FE-HMM solution for problem (80) for decreasing macro mesh $H_M = 1/4, 1/8, 1/16$ (optimal micro mesh chosen accordingly).

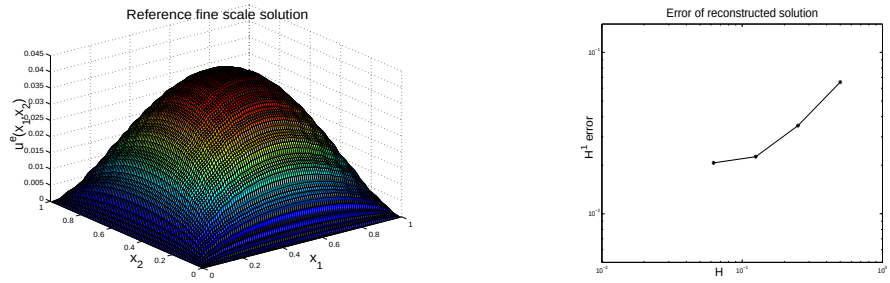


Figure 20: Reference solution for problem (80) (left picture) and convergence rate (broken H^1 norm) with adapted (optimal) micro mesh refinement (right picture).

Finally, we present in Figure 21 the FE-HMM fine scale reconstructed solutions on a sampling domain $K_\varepsilon(x_K)$. The snapshot (taken on the same sampling domain) is taken from

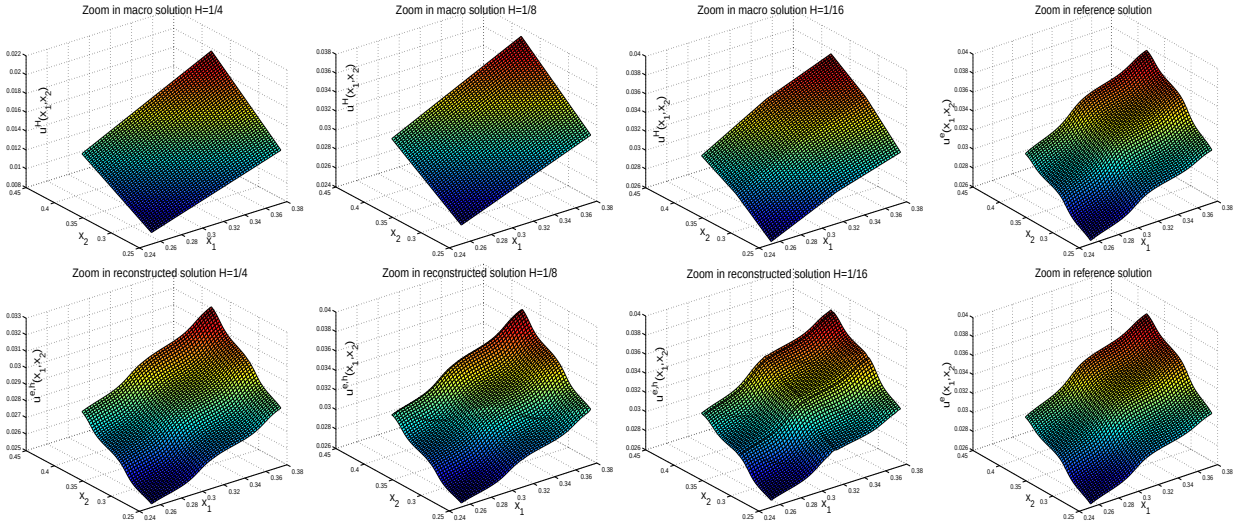


Figure 21: Zoom in the FE-HMM solution on sampling domains (first row of pictures) and econstructed (smal-scale) FE-HMM solution on sampling domains (second row of pictures) The various pictures (in each row) correspond to the various macro-meshes $H_M = 1/4, 1/8, 1/16$ of Figure 16. The last picture (in each row) is a zoom of the reference small scale solution for the given sampling domain.

several computations with successively refined macro meshes. We can see that while the FE-HMM solution cannot capture the oscillation of the fine scale solution, this oscillations can be captured by the reconstructed solution, even with a very coarse macro mesh.

4. HMM based on discontinuous Galerkin FE and spectral FE

As should be clear from the discussion in the previous sections, the framework of the HMM allows for flexibility in coupling macro and micro methods. The coupling conditions are not strictly enforced and besides periodic and Dirichlet boundary conditions presented previously, Neumann or Robin boundary conditions for example could also be used. Another flexibility in the FE-HMM methodology is in the choice of the macro and micro FE spaces. Over the years, an impressive body of FE methods have been developed for various classes of applications, as mixed FEM, discontinuous Galerkin FEM, mortar FEM, partition of unity FEM, spectral FEM to mention but a few. The FE-HMM can potentially accommodate such methods at the macro or the micro scale, although specific modeling and analysis issues depending on the chosen method have to be addressed and may not be trivial. The question in a macro-micro framework is thus: what are the desired properties (which may of course depend on the application) at a given scale and how can we couple different methods to match these properties ? In the following we briefly discuss two recent developments in the direction of such "qualitative coupling" or "hybrid methods".

4.1. Finite element heterogeneous multiscale methods with near optimal computational complexity. In order to reduce the overall super-linear complexity (in the macro DOF) of the FE-HMM, a method coupling FE (macro solver) and spectral methods (micro solver), the so-called FES-HMM, has been proposed in [10]. Provided sufficient regularity of the conductivity tensor, the micro solution in the FES-HMM has *spectral accuracy* or even

exponential convergence, and the overall complexity is quasi optimal, i.e. *almost-linear* in the macro (N_{mac}) DOF.

The idea of the method is the following. We consider a modified bilinear form as defined in (31), but where the micro functions, that we denote here as $v_{j,M}, w_{j,M}$, are the solution of the following problem: for $u^H \in V^p(\Omega, \mathcal{T}_H)$ find u_M such that $(u_M - u^H_{lin,j}) = w_M \in S_M(K_{\delta_j})$ and

$$(a^\varepsilon \nabla w_M, \nabla z_M)_M = (a^\varepsilon \nabla u^H_{lin,j}, \nabla z_M)_M, \ \forall z_M \in S_M(K_{\delta_j}), \tag{89}$$

where

$$S_M(K_{\delta_j}) := \text{span}\{e^{2i\pi kx/\varepsilon}; x \in K_{\delta_j}, k \in \mathbb{Z}^d, -M \le k_i \le M-1\}/\mathbb{R}. \tag{90}$$

Given a mesh $\{\xi_l\}_{l_1,\dots,l_d=0}^{2M-1}$ on the sampling domain K_{δ_j}, $(\cdot,\cdot)_M$ denotes a discrete scalar product given by

$$(u,v)_M := \frac{|K_{\delta_j}|}{(2M)^d} \sum_{l_1,\dots,l_d=0}^{2M-1} u(\xi_l)\bar{v}(\xi_l). \tag{91}$$

Spectral methods are particularly powerful on simple geometries and this can be exploited in the FES-HMM, since it is the macro triangulation which meshes the physical domain and the sampling domains are usually chosen as squares or cubes. A fully-discrete analysis of the FES-HMM has been obtained in [10], where numerical examples also with non-periodic coefficients (as the problem with random coefficients (15)) have also been presented. It is shown in [10] that up to spectral or exponential convergence of the micro FEM, the overall complexity of the method is *near optimal*, i.e., $\mathcal{O}(N_{mac})$ floating point operations for a linear (macro) convergence rate in the H^1 norm and $\mathcal{O}(N_{mac})$ floating point operations for a quadratic convergence rate in the L^2 norm (compare these results with the complexity of the FE-HMM discussed in Section 3).

As an illustration, let us consider the problem (79), this time solved with the FES-HMM. The parameters and the notation for this example are the same as chosen in Section 3.

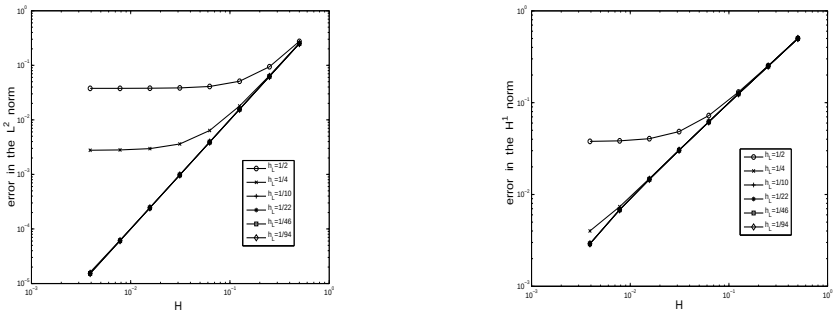


Figure 22: Error against the homogenized solution u^0 for the FES-HMM applied to problem (79) with macromesh refinement for fixed micro mesh $h_L = 1/2, 1/4, 1/10, 1/22, 1/46$.

Numerical results are presented in Figure 22. We see that provided $h_L \ge 1/8$, the error in the numerical results computed in both the L^2 and the H^1 norms are no longer dependent on the small scale, i.e., up to spectral convergence, the error of the FES-HMM is given by the usual error of the macro FEM, and the complexity only depends on the macro DOF. A

comparison with Figure 13, where micro and macro meshes must be simultaneously refined during the whole macro mesh refinement process, shows the remarkable improvement.

4.2. Discontinuous Galerkin finite element heterogeneous multiscale method (DG-HMM) For many problems, local conservation properties in the numerical approximation and flexibility in meshing (e.g. handling hanging nodes, performing local refinement) are desirable. A popular methodology to achieve these aims is to use discontinuous (local) FE approximations. For multiscale problems and in a macro-micro methodology as the FE-HMM, the aforementioned properties are primarily desirable at the macro level, on the physical domain. We will here focus on the discontinuous Galerkin (DG) methods. Such methods have been extensively studied for hyperbolic problems, advection-diffusion and diffusion problems (see [18] and the references therein). In [11], the first analysis of a multiscale DG methods for elliptic homogenization problems has been given. Multiscale methods for one-dimensional hyperbolic and parabolic problems have been proposed in [23] in the HMM framework. For elliptic problems, a DG-FEM has recently been proposed in [14] for homogenization problems, however, not in an HMM framework and without analysis.

We briefly describe here the DG-HMM given in [11]. In a DG framework, we relax the standard interelement continuity for FEM and we consider the FE space (given here for piecewise linear discontinuous FE)

$$V^1_{DG}(\Omega, \mathcal{T}_h) \;=\; \{u^h \in L^2(\Omega);\ u^h|_K \in \mathcal{P}^1(K),\ \forall K \in \mathcal{T}_H\}. \tag{92}$$

Notice that requiring only that $u^h \in L^2(\Omega)$ does not ensure continuity of u^h at the interfaces of elements where these functions can have jumps. Many types of DG-FEM have been developed (see [18]) and we only briefly describe in what follows the so-called interior penalty DG-FEM. We consider an arbitrary element K of our triangulation $\mathcal{T}_h$, multiply the problem (16) with a smooth test function v and integrate by parts using $a^\varepsilon \nabla u^\varepsilon \in H(\mathrm{div}, K)$. Summing over $K \in \mathcal{T}_h$ yields

$$\int_\Omega a^\varepsilon \nabla u^\varepsilon \cdot \nabla v dx - \sum_{K\in\mathcal{T}_h} \int_{\partial K} a^\varepsilon \nabla u^\varepsilon \cdot n_K v ds = \int_\Omega f v dx, \tag{93}$$

where n_K is the outward normal. We denote by $e \in \mathcal{E}$ an interface shared by two neighboring elements K_1 and K_2, where $\mathcal{E}$ is the set of all (interior and boundary) interfaces. Since hanging nodes are allowed, $\mathcal{E}$ will be understood to contain the smallest common interfaces of neighboring elements. For a piecewise smooth function ξ (possibly vector valued) denote by ξ_1, ξ_2 its trace from within K_1, K_2, respectively, and the average and the jump defined by $\{\xi\} = \frac{1}{2}(\xi_1 + \xi_2)$, $\quad [\![\xi]\!] = \xi_1 n_1 + \xi_2 n_2$, respectively, where n_i denotes the unit outward normal vectors on the interface K_i. Notice that $[\![\xi]\!]$ is a vector-valued function if ξ is a scalar function, while it is a scalar function if ξ is a vector-valued function. Using these notations we can rewrite (93) as

$$\int_\Omega a^\varepsilon \nabla u^\varepsilon \cdot \nabla v dx - \sum_{e\in\mathcal{E}} \int_e \{a^\varepsilon \nabla u^\varepsilon\}\ [\![v]\!] = \int_\Omega f v dx. \tag{94}$$

Since the exact solution of (16) is in $H^1_0(\Omega)$ we have $[\![u^\varepsilon\ ]\!] = 0$ and we can make the bilinear form (94) symmetric by adding $-\sum_{e\in\mathcal{E}} \int_e \{a^\varepsilon \nabla v\}\ [\![u]\!]$ (assuming the existence of a trace for

$a^\varepsilon \nabla v$). Finally, to have a stable method one adds a penalty term. All together we obtain the interior penalty DG-FEM (see [18]) for which one seeks a solution $u^h \in V_h$ such that

$$\int_\Omega a^\varepsilon \nabla u^h \cdot \nabla v^h dx - \sum_{e\in\mathcal{E}} \int_e \left(\{a^\varepsilon \nabla u^h\}\, [\![v^h]\!] + \{a^\varepsilon \nabla v^h\}\, [\![u^h]\!]\right) ds + \sum_{e\in\mathcal{E}} \int_e \mu [\![u^h]\!][\![u^h]\!] = \int_\Omega f v^h dx \ \forall v^h \in V_h, \tag{95}$$

where $\mu = \alpha h_e^{-1}$ with $\alpha > 0$ independent of the meshsizes and h_e is the interface size (with the above convention for hanging nodes). Here and in what follows, the gradient ∇ should be understood as a broken gradient ∇_h when dealing with discontinuous functions defined by $\nabla_h u^h|_K = \nabla u,\ \forall K \in \mathcal{T}_h$. The choice of α is dictated by stability requirement. The analysis of this method as well as many other methods based on discontinuous Galerkin FE space is discussed in [18].

Let us make a few remarks. First, as for FEM $h < \varepsilon$ is required to have a good approximation and this is prohibitive in terms of computation costs if ε is small. Second, regularity on a^ε to be able to extend it up to ∂K is needed and this may not be realistic for many problems with oscillating coefficients. In the method described below, we will only need to compute averages of a^ε on sampling domains and we will thus not require the existence of traces for a^ε.

The goal is now to define a modified bilinear form similar to (31) (given in what follows for piecewise linear polynomial) but based on the macro DG space

$$V^1_{DG}(\Omega, \mathcal{T}_H) \ = \ \{u^H \in L^2(\Omega);\ u^H|_K \in \mathcal{P}^1(K),\ \forall K \in \mathcal{T}_H\},$$

where H is allowed to be much larger than ε.

The DG-HMM. For $v^H, w^H \in V^1(\Omega, \mathcal{T}_H)$ we define $B_{DG}(u^H, v^H) =$

$$\sum_{K\in\mathcal{T}_H} \omega_{K_\delta} \int_{K_\delta} a^\varepsilon \nabla u^h \cdot \nabla v^h dx - \sum_{e\in\mathcal{E}} \int_e \left(\{\overline{a^\varepsilon \nabla u^h}\}[\![v^H]\!] + \{\overline{a^\varepsilon \nabla v^h}\}[\![u^H]\!]\right) ds \tag{96}$$
$$+ \sum_{e\in\mathcal{E}} \int_e \mu [\![u^H]\!][\![v^H]\!],$$

where μ is the discontinuity-penalization parameter defined by $\mu|_e = \mu_e = \alpha H_e^{-1}$ (with the same convention as before for hanging nodes) and α is a positive parameter independent of the local meshsize. The micro functions are defined similarly as for the FE-HMM (see (31)). The important modeling issue is now to define appropriate multiscale flux averages $\{\overline{a^\varepsilon \nabla v^h}\}, \{\overline{a^\varepsilon \nabla w^h}\}$. This can be done as follows.

For an interior interface e of two triangles $K_i\ i = 1, 2$ with sampling domains $K_{\delta,i}\ i = 1, 2$ and a boundary interface of a triangle K with sampling domain K_δ we define

$$\{\overline{\xi}\} = \frac{1}{2}\left(\frac{1}{|K_{\delta,1}|}\int\limits_{K_{\delta,1}} \xi_1 dx + \frac{1}{|K_{\delta,2}|}\int\limits_{K_{\delta,2}} \xi_2 dx\right),\ \{\overline{\xi}\} = \left(\frac{1}{|K_\delta|}\int\limits_{K_\delta} \xi_1 dx\right),$$

respectively, where ξ is an integrable (possibly vector valued) function.

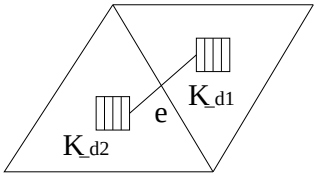


Figure 23: Illustration of the modeling of the multiscale flux average. The flux average for two macro triangles K_1, K_2 with sampling domains $K_{\delta_1}, K_{\delta_2}$ for an interface e is based upon an average of the micro flux computed in each sampling domain.

The macro solution of the DG-HMM is then defined by the following variational problem: find $u^H_{DG} \in V^1_{DG}(\Omega, \mathcal{T}_H)$ such that

$$B_{DG}(u^H_{DG}, v^H) = \int_\Omega f v^H dx, \quad \forall v^H \in V^1_{DG}(\Omega, \mathcal{T}_H). \tag{97}$$

Several remarks are in order. First, the computational saving compared to (95) for a multiscale problem (16) is clear since instead of solving the fine scale on the whole computational domains (as required for (95) with $h < \varepsilon$), in the DG-HMM we only solve the fine scale on sampling domains K_δ usually of much smaller size than the macro meshsize H. Second, we do not require well-defined traces of a^ε on ∂K as was needed in (95). Third, the interface contribution are based on macro functions and averaged micro fluxes already available from the computation of the first term of (96). Fourth, the method is designed for coefficients a^ε of general type. Error estimates, obtained for non-uniformly periodic coefficients, and details about the method can be found in [11].

5. Conclusion and perspectives

In this paper, we have discussed in details the modeling and the analysis of a multiscale method, the FE-HMM, for homogenization problems. We have shown that the method is flexible enough to allow for different types of discretizations and different types of problems. Besides elliptic problems, we have presented numerical examples for advection-diffusion and non-linear parabolic problems. We note that, although not discussed in this paper, elasticity problems have been treated in [8]. We have also shown that the framework used to construct the FE-HMM allows for precise convergence analysis and give in turn a criteria for mesh refinement. The methodology used here has yet another nice property: it allows for simple coding, since the structure of standard FEM can be used at the macro level. We did not discuss the implementation issues which are reported in [12], where a short code (allowing to compute all the examples presented in Section 2) is given. Also a generalization of the FE-HMM for more than two scales can be done but need yet to be analyzed. A crucial assumption as mentioned throughout this paper, is that of scale separation. This is realistic for many applications although sometimes only in some region of the computational domain and/or for some period of time. The importance of constructing adaptive methods can thus not be overemphasized. The boundary layers and coupling issues arising when using sampling domains which do not match the small scale oscillations, need also to be better understood. Adaptivity and robustness are thus central issues which need to be addressed in future research.

Acknowledgments. The author thanks A. Nonnenmacher for a careful reading of the manuscript. This work is partially supported by an EPSRC Advanced Fellowship EP/E05207X/1.

References

[1] A. Abdulle, *Fourth order Chebyshev methods with recurrence relation*, SIAM J. Sci. Comput., 23, 6 (2002), pp. 2041-2054.

[2] A. Abdulle and W. E, *Finite Difference HMM for homogenization problems*, J. Comput. Phys., 191 (2003), pp. 18-39.

[3] A. Abdulle and S. Attinger *Homogenization method for transport of DNA particles in heterogeneous arrays,* Multiscale Model. and Simul., Lect. Notes., Springer, 39 (2004), pp. 23-33.

[4] A. Abdulle and S. Attinger *Numerical methods for transport problems in micro devices,* Springer Lecture Notes in Comput. Sci., 3743, Spriner Berlin (2006), 67-75.

[5] A. Abdulle and C. Schwab, *Heterogeneous Multiscale FEM for Diffusion Problem on Rough Surfaces* , SIAM Multiscale Model. Simul., 3, 1 (2005), pp. 195-220.

[6] A. Abdulle, *On a-priori error analysis of Fully Discrete Heterogeneous Multiscale FEM*, SIAM Multiscale Model. Simul., 4, 2 (2005), pp. 447-459.

[7] A. Abdulle, *Multiscale methods for advection-diffusion problems,* Discrete and Contin. Dyn. Syst., suppl. (2005), pp. 11-21.

[8] A. Abdulle, *Analysis of a Heterogeneous Multiscale FEM for Problems in Elasticity*, Math. Mod. Meth. Appl. Sci. (M3AS), 16, 2 (2006), pp. 1-21.

[9] A. Abdulle, *Heterogeneous multiscale methods with quadrilateral finite elements,* Numerical mathematics and advanced applications, Springer, Berlin, (2006), pp. 743-751.

[10] A. Abdulle and B. Engquist, *Finite element heterogeneous multiscale methods with near optimal computational complexity,* SIAM Multiscale Model. Simul., 6 (2007), pp. 1059-1084.

[11] A. Abdulle, *Multiscale method based on discontinuous finite element methods for homogenization problems*, C. R. Acad. Sci. Paris, Ser. I, 346 (2007), pp. 97-102.

[12] A. Abdulle and A. Nonnenmacher, *A short and versatile finite element multiscale code for homogenization problems*, to appear in CMAME 2009.

[13] A. Abdulle and P. Heuser, *Homogenization and multiscale method for quasilinear elliptic-parabolic equations with application to Richards equation,* in preparation.

[14] J. Aarnes and B.-O. Heimsund, *Multiscale discontinuous Galerkin methods for elliptic problems with multiple scales,* Lecture Notes in Comp. Sci. and Eng., 44, (2006), pp. 1-20.

[15] G. Allaire and R. Brizzi, *A multiscale finite element method for numerical homogenization*, SIAM Multiscale Model. Simul., 4 (2005), pp. 790-812.

[16] S. Attinger and A. Abdulle *Effective velocity for transport in heterogeneous compressible flows with mean drift,* Phys. Fluids 20, 016102 (2008).

[17] P.M. Adler, J.-F. Thovert, S. Berki and F. Yousifian, *Real Porous Media: Local Geometry and Transports*, J. of Engineering Mechanics, 128, 8 (2002), pp. 829-839.

[18] D. Arnold, F Brezzi, B. Cockburn and D. Marini, *Unified Analysis of discontinuous Galerkin Methods for elliptic problems*, Siam J Numer. Anal., 5 (2002), pp. 1749–1779.

[19] I. Babuska, *Homogenization and its applications*, SYNSPADE 1975, B. Hubbard ed., pp. 89-116.

[20] T.J. Barth, T. Chan and R. Haimes, eds. "Multiscale and multiresolution methods, theory and application", Springer, Lecture Notes in Computational Science and Engineering, 20, 2002.

[21] R. Bausch and R. Schmitz, *Diffusion in the Presence of Topological Disorder*, Physical Review Letters, 73, 18 (1994), pp. 2382–2385.

[22] A. Bensoussan, J.-L. Lions and G. Papanicolaou, "Asymptotic Analysis for Periodic Structure", North Holland, Amsterdam, 1978.

[23] S. Chen, W. E and C. Shu,*The Heterogeneous Multiscale Method Based on the Discontinuous Galerkin Method for Hyperbolic and Parabolic Problems,* SIAM Multiscale Model. Simul., 3, 4 (2005), pp. 871-894.

[24] P.G. Ciarlet and P.A. Raviart, *The combined effect of curved boundaries and numerical integration in isoparametric finite element methds*, in The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations, A. K. Aziz, ed., Academic Press, New York, (1972), pp. 409-474.

[25] P.G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North-Holland, 1978.

[26] D. Cioranescu and P. Donato, "An introduction to Homogenization", Oxford University Press, 1999.

[27] Z. Chen, W. Deng and H. Ye, *Upscaling of a class of nonlinear parabolic equations for the flow transport in heterogeneous porous media,* Comm. Math. Sci., 3, 4 (2005), pp. 493-515.

[28] T.A. Duke and R.H. Austin, *Microfabricated Sieve for the Continuous Sorting of Macromolecules* Phys. Rev. Lett. 89, 7 (1998).

[29] M. Dorobantu, B. Engquist, *Wavelets-based numerical homogenization*, SIAM J. Numer. Anal., 35, 2 (1998), pp. 540–559.

[30] B. Engquist, *Computation of oscillatory solutions to hyperbolic differential equations*, Springer Lecture Notes in Mathematics, 1270, (1987), pp.10-22.

[31] B. Engquist and O. Runborg, *Wavelets-Based Numerical Homogenization with Applications*, Multiscale and Multiresolution Methods, Lecture Notes in Computational Science and Engineering, 20, Springer Verlag, (2002), pp. 97-148.

[32] W. E and B. Engquist, *The Heterogeneous Multi-Scale Methods*, Commun. Math. Sci., 1 (2003), pp. 87-132.

[33] W. E, P. Ming and P. Zhang, *Analysis of the heterogeneous multiscale method for elliptic homogenization problems*, J. of AMS, 18, 1 (2004), pp. 121-156.

[34] W. E, B. Engquist, X. Li, W. Ren, E. Vanden-Eijden, *The heterogeneous multiscale method: A review*, Comm. in Comput. Physics, 2 (2007), pp. 367-450.

[35] Y. Efendiev, T. Hou and V. Gitting, *Multiscale finite element method for nonlinear problems and their applications,* Comm. Math. Sci., 2, 4 (2004), pp. 553-589.

[36] D. Ertas, *Lateral Separation of Macromolecules and Polyelectrolytes in Microlithographic Arrays* Phys. Rev. Lett.,80, 7 (1998).

[37] M.T. van Genuchten, *A closed-form equation for predicting the hydraulic conductivity of unsaturated soils,* Soil. Sci. Soc. Am. J., 44, (1980), pp. 892-898.

[38] E. De Giorgi and S. Spagnolo, *Sulla convergenza degli integrali dell'energia per operatori ellittici del secondo ordine* Boll. UMI, 4, 8 (1973), pp. 391-411.

[39] D. Givon, R. Kupferman and A.M Stuart, *Extracting macroscopic dynamics: model problems and algorithms,* Nonlinearity, 17 (2004), pp. 55-127.

[40] E. Dimmler, R. Marabini, P. Tittmann and H. Gross, *Correlation of Topographic Surface and Volume Data from Three-Dimensional Electron Microscopy*, Journal of Structural Biology, 136, (2001), pp. 20-29.

[41] L.R. Huang, P. Silberzan, J.O. Tegenfeldt, E.C. Cox, J.C. Sturm, R.H. Austin and H. Craighead, *Role of Molecular Size in Ratchet Fractionation*, Phys. rev. Lett., 89, 17 (2002).

[42] T-Y. Hou, X-H. Wu and Z. Cai, *Convergence of a multi-scale finite element method for elliptic problems with rapidly oscillating coefficients*, Math. of Comput., 68, 227 (1999), pp. 913-943.

[43] P. Heuser, *Homogenization of quasilinear elliptic-parabolic equations with application to Richards equation*, Preprint No. 2006-10, Departement Mathematik, University of Basel, October 2006.

[44] Viet Ha Hoang and Christoph Schwab, *High-Dimensional Finite Elements for Elliptic Problems with Multiple Scales*, SIAM Multiscale Model. Simul., 3, 1 (2005), pp. 195-220.

[45] V.V. Jikov, S.M. Kozlov and O.A. Oleinik, *Homogenization of differential Operators and Integral Functionals"*, Springer-Verlag, Berlin, Heidelberg, 1994.

[46] T. Fujiwara, K. Ritchie, H. Murakoshi, K. Jacobson and A. Kusumi, *Phospholipids undergo hop diffusion in compartmentalized cell membrane*, Journal of Cell Biology, 157, 6 (2002), pp.1071-1081.

[47] K. Kordás et al., *Chip cooling with integrated carbon nanotube microfin architectures,* Appl. Phys. Lett., 90, 123105 (2007).

[48] V. Kouznetsova, W.A.M. Brekelmans and F.P.T. Baaijens, *An approach to micro-macro modeling of heterogeneous materials*, Computational Mechanics, 27, (2001), pp. 37-48.

[49] A. J. Majda and P. R. Kramer, *Simplified models for turbulent diffusion: Theory, numerical modelling and physical phenomena,* Physics Reports, 314, (1999), pp. 237–574.

[50] A. Mikelić, C. Rosier, *Modeling solute transport through unsaturated porous media using homogenization*, Comp. and Appl. Math., 23, 2-3 (2004), pp. 195–211.

[51] C. Miehe, J. Schröder and C. Bayreuther, *On the homogenization analysis of composite materials based on discretized fluctuations on the micro-structure*, Acta Mechanica, 135, (2002), pp. 1-16.

[52] A-M. Matache, I. Babuska and C. Schwab, *Generalized p-FEM in Homogenization*, Numerische Mathematik, 86, 2 (2000), pp. 319-375.

[53] A.M. Matache and C. Schwab, *Two-scale FEM for Homogenization Problems*, R.A.I.R.O. Anal. Numerique, 36 (2002), pp. 537-572.

[54] P. Ming and Y. Yue, *Numerical methods for multiscale elliptic problems,* J. Comput. Physics, 214, 1 (2006), pp. 421-445.

[55] P. Ming and P. Zhang, *Analysis of the herogeneous multiscale method for parabolic homogenization problems*, Math. Comp., 76, 257 (2007), pp. 153-177.

[56] V.V. Mourzenko, J.-F. Thovert and P.M. Adler, *Percolation and conductivity of self-affine fractures*, Physical Review E, 59, 4 (1999), pp. 4265-4284.

[57] F. Murat and L. Tartar *H-convergence,* in "Topics in the mathematical modeling of composite materials", A. Cherkaev and R. Kohn Eds., Birkhäuser, Boston, (1997), pp. 21-43.

[58] N. Neuss, W. Jäger and G. Wittum, *Homogenization and multigrid*, Computing, 66, 1 (2001), pp. 1-26.

[59] J.T. Oden and K.S. Vemaganti, *Estimation of local modeling error and global-oriented adaptive modeling of heterogeneous materials: error estimates and adaptive algorithms,* J. Comput. Phys., 164, 1 (2000), pp. 22-47.

[60] G. Pavliotis, *Homogenization Theory for Advection–Diffusion Equations with Mean Flow*, PhD Thesis, Rensselaer Polytechnic Institute, May 2002.

[61] I.S. Pop and W.A. Yong, *A numerical approach to degenerate parabolic equations,* Numer. Math., 92 (2002), pp. 357-381.

[62] M. Slodička, *A robust and efficient linearization scheme for doubly nonlinear and degenerate arabolic problems arising in flow in porous media*, SIAM J. Sci. Comput., 23, 5 (2002), pp. 1593–1614.

[63] K. Terada and N. Kikuchi, *A class of general algorithm for multi-scale analyses of heterogeneous media,* Comput. Methods Appl. Mech. Engrg.,190 (2001), pp. 5427–5464.

[64] T.C. Wallstrom, S. Hou, M.A. Christie, L.J. Durlofsky, D.H. Sharp, *Accurate scale up of two phase flow using renormalization and nonuniform coarsening*, Computational Geosciences, 3, 1 (1999), pp. 69-87.

[65] X. Yue and W. E, *The local microscale problem in the multiscale modeling of strongly heterogeneous media: Effects of boundary conditions and cell size*, J. Comput. Phys., 222, 2 (2007), pp. 556–572.

Assyr Abdulle
School of Mathematics and Maxwell Institute for Mathematical Sciences
University of Edinburgh
JCMB, King's Buildings
Edinburgh EH9 3JZ
UK
e-mail: a.abdulle@ed.ac.uk

Present address:
École Polytechnique Fédérale de Lausanne (EPFL)
Section de Mathématiques SB-SMA
Station 8 - Bâtiment MA, 1015 Lausanne
Switzerland
e-mail: assyr.abdulle@epfl.ch